

Apprentissage par réseau de neurones pour communiquer en présence d'une interférence radar

Florence ALBERGE

Laboratoire des Signaux et Systèmes (L2S)
Université Paris-Sud, CNRS, CentraleSupélec, Université Paris-Saclay
3, rue Joliot-Curie, 91192 Gif-sur-Yvette, France
florence.alberge@u-psud.fr

Résumé – Nous nous intéressons à l'utilisation partagée du spectre entre un radar et un système de communication sans fil. C'est une des pistes envisagée pour lutter contre la congestion des réseaux de communication. Le canal de transmission est modélisé par la somme d'un bruit gaussien et d'une interférence radar. Les constellations standards ne sont pas efficaces dans ce contexte. Nous proposons ici d'utiliser une technique d'apprentissage par réseaux de neurones pour le design de la constellation et du détecteur. Les méthodes d'apprentissage sont particulièrement appropriées lorsque l'on ne dispose pas d'une expression analytique de la fonction de coût ou lorsque celle-ci est trop compliquée. La méthode proposée améliore significativement les performances par rapport à la configuration standard. Elle est valide quel que soit le niveau d'interférence contrairement aux méthodes proposées jusqu'ici dans la littérature.

Abstract – Radar and wireless communication coexistence is considered in this paper as a possible solution to face the exploding demand and rising congestion in wireless networks. The transmission medium is modeled as an AWGN channel with additive radar interference. Standard constellations are not optimal in this context and deep learning is used to design proper constellations and corresponding receiver devices. Deep learning is particularly relevant in the lack of an analytical expression of the loss function. The proposed method is able to yield solutions that outperform the standard configurations. This is a step further compared to previous works in which solutions were limited to low or high interference regimes.

1 Introduction

L'augmentation du trafic sur les réseaux sans fil oblige à repenser la politique d'allocation des ressources. Allouer des bandes de fréquence à des usages ou des services spécifiques est sous-optimal. Certaines portions du spectre, comme celles réservées aux radars et infrastructures de radio-navigation, sont sous-utilisées. Plusieurs régulateurs envisagent maintenant d'en autoriser l'accès à des services commerciaux. De nombreux travaux de recherche s'intéressent donc à cette situation d'utilisation partagée du spectre [1]. Une partie importante de ces travaux s'est concentrée sur le design du système radar et du système de communication afin d'en réduire l'impact sur l'autre dispositif. Nous nous intéressons ici à une approche plus opportuniste dans laquelle le radar pré-existe et où le système de communication doit s'adapter pour lutter au mieux contre l'interférence du système radar. La capacité d'un canal de communication avec bruit additif gaussien et interférence radar a été étudiée dans [2]. Il y est prouvé que ce canal permet une communication quelque soit le niveau d'interférence. Les cas les plus sévères sont observés lorsque l'interférence et le message transmis ont des ordres de grandeur comparables. L'objectif de cet article est de trouver le détecteur et la constellation optimale pour ce type de canal. Cette question a été également considérée dans [3]. Des formules analytiques du taux d'erreur

ont été données pour les deux cas extrêmes (faible et forte interférence) mais ils ne se généralisent pas aux situations intermédiaires. Nous proposons ici une approche radicalement différente qui conduit à un cadre global applicable pour tout niveau d'interférence. Un détecteur quasi-optimal est proposé, il englobe comme cas particulier les détecteurs proposés dans la littérature pour les cas de faible et forte interférence. La constellation est optimisée à l'aide d'une technique d'apprentissage par réseaux de neurones. Elle fournit, en plus de la constellation, un détecteur que nous comparerons au détecteur précédent. Nous montrons enfin par simulation numérique le gain apporté par la méthode proposée.

2 Modèle et notations

Nous nous intéressons dans cet article à un modèle de communication dans lequel l'interférence est principalement due à un radar bande étroite et à faible rapport cyclique. Nous considérons un canal mono-porteuse, lent et plat en fréquence [4, 2]. Le signal reçu s'écrit alors

$$Y = \sqrt{S}X + \sqrt{I}e^{j\theta_I} + Z \quad (1)$$

où Z est un bruit gaussien de moyenne nulle et de variance 1 et X est l'entrée du canal, X est un point de la constellation

\mathcal{X} . Les variables aléatoires (X, θ_I, Z) seront supposées indépendantes les unes des autres. Avec ces notations, le rapport signal sur bruit (SNR) est égal à S et le rapport interférence sur bruit (INR) est égal à I . Ces deux quantités sont supposées connues. L'objectif est de trouver la meilleure constellation \mathcal{X} et le meilleur récepteur pour le canal à interférences donné dans (1). Des résultats ont déjà été obtenus dans les deux cas limite (faible et forte interférence) [2, 3]. Nous proposons ici une méthode plus générale applicable quelle que soit la valeur de l'interférence. Le modèle de canal est simple mais les résultats peuvent être étendus à un canal variant dans le temps en moyennant les taux d'erreurs à l'aide de la statistique du canal. Ces résultats pourront en outre servir de point de départ à une généralisation aux systèmes multi-porteuses.

3 Constellation optimale

Pour choisir la constellation à transmettre, nous utilisons une technique d'apprentissage à base de réseau de neurones. Un réseau de neurones est composé d'une succession de couches, les sorties de la couche précédente servant d'entrée à la couche suivante. Pour une couche donnée, si l'on note ℓ_{in} le signal en entrée et ℓ_{out} le signal en sortie, on a

$$\ell_{out} = g(\mathbf{w}\ell_{in} + \mathbf{b}) \quad (2)$$

où \mathbf{w} et \mathbf{b} sont les paramètres qui seront appris lors de l'entraînement sur une base d'exemples. Les paramètres sont optimisés conjointement pour l'ensemble des couches à partir des données d'entraînement par rétropropagation du gradient de l'erreur [5]. Soit d_{in} la longueur du vecteur d'entrée ℓ_{in} et d_{out} la longueur du vecteur de sortie ℓ_{out} , la matrice \mathbf{w} a donc pour dimension $d_{out} \times d_{in}$ et le vecteur \mathbf{b} a pour longueur d_{out} . La fonction g est appelée fonction d'activation. elle s'applique à chaque élément du vecteur $\mathbf{w}\ell_{in} + \mathbf{b}$. Dans la suite, nous utiliserons la notation compacte

$$\ell_{out} = f_{\mathcal{P}}^{d_{in} \rightarrow d_{out}, g}(\ell_{in}) \quad (3)$$

où $\mathcal{P} = \{\mathbf{w}, \mathbf{b}\}$. Le réseau de neurones utilisé ici est donné sur la figure 1. Il se compose de deux parties. Le premier réseau de neurones implémente l'association entre un mot binaire C de longueur k et un point X de la constellation \mathcal{X} . Cette opération est réalisée par un réseau de neurones à 4 couches selon :

$$X = f_{\mathcal{P}_4}^{4 \rightarrow 2, l} \circ f_{\mathcal{P}_3}^{8 \rightarrow 4, t} \circ f_{\mathcal{P}_2}^{16 \rightarrow 8, t} \circ f_{\mathcal{P}_1}^{k \rightarrow 16, t}(C) \quad (4)$$

où t signifie que la fonction d'activation est la tangente hyperbolique alors que l indique une activation linéaire. Le rôle de ce premier réseau de neurones est donc de trouver la constellation \mathcal{X} la mieux adaptée au canal et de déterminer la meilleure association entre les mots binaires et les points de la constellation. La normalisation de l'énergie garantit la condition $E[|X|^2] = 1$. Dans les cas où l'interférence n'est pas prépondérante, le canal est celui donné dans (1). Lorsque l'interférence est prépondérante on peut estimer θ_I par $\hat{\theta}_I = \arg(Y)$. On note $Y_{IC} = Y - \sqrt{I}e^{j\hat{\theta}_I}$ la sortie du canal après suppression de l'interférence estimée. Le deuxième réseau de neurones implémente le

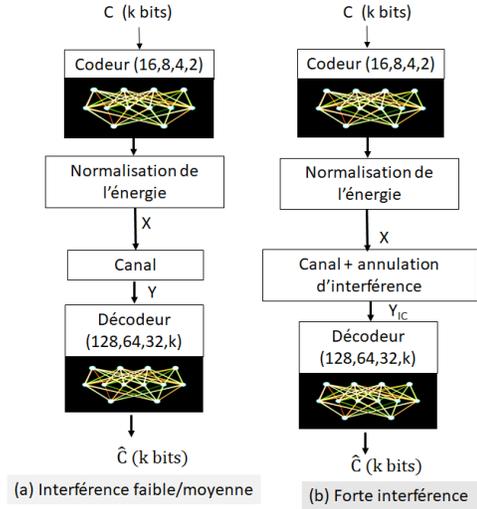


FIGURE 1 – Schéma du réseau de neurones.

détecteur, il a donc pour entrée les observations en sortie du canal (Y ou Y_{IC}). Il calcule une estimation \hat{C} du mot binaire C . Ce décodeur contient également 4 couches choisies de la manière suivante :

$$\hat{C} = f_{\mathcal{P}'_4}^{32 \rightarrow k, s} \circ f_{\mathcal{P}'_3}^{64 \rightarrow 32, r} \circ f_{\mathcal{P}'_2}^{128 \rightarrow 64, r} \circ f_{\mathcal{P}'_1}^{2 \rightarrow 128, r}(Y) \quad (5)$$

où r signifie que la fonction d'activation est $r(x) = \max(0, x)$ et où s fait référence à la fonction sigmoïde. Le détecteur est similaire à celui utilisé par [6] dans un autre contexte. Pour une session d'entraînement, le triplet (k, S, I) est fixé. L'ensemble d'apprentissage est constitué des 2^k mots binaires possibles. Pendant la phase d'entraînement $\hat{C} = C$. Le réseau est entraîné afin de minimiser l'erreur de reconstruction $-\log p(C|\hat{C})$. On utilise pour cela un gradient stochastique avec un pas fixe $l_r = 0.01$. La taille des *batches* est 2^k et le nombre d'*epochs* est 2^{19} [7]. L'apprentissage est réalisé avec Keras [8]. Une fois l'entraînement réalisé l'équation (4) donne la correspondance entre les mots binaires et les points de la constellation. L'équation (5) donne le détecteur. Dans la section suivante, nous proposons un deuxième détecteur par approximation du détecteur optimal.

4 Détecteur optimal

A partir de l'équation (1), on peut déterminer la probabilité conditionnelle $p_{Y|X, \theta}$ selon

$$p_{Y|X, \theta}(y, x|\theta) = \frac{1}{\pi} e^{-|y - \sqrt{S}x - \sqrt{I}e^{j\theta_I}|^2} \quad (6)$$

où $y, x \in \mathbb{C}$ and $\theta_I \in [0, 2\pi)$. Comme θ_I est inconnu, nous considérons la moyenne sur la variable θ_I soit

$$p_{Y|X}(y, x) = \frac{1}{2\pi} \int_0^{2\pi} p_{Y|X, \theta_I}(y, x|\theta) d\theta_I \quad (7)$$

$$= \frac{e^{-|y - \sqrt{S}x|^2 - I}}{\pi} I_0(2\sqrt{I}|y - \sqrt{S}x|) \quad (8)$$

où I_0 est la fonction de Bessel modifiée de première espèce et d'ordre 0 [2]. Le détecteur optimal au sens du maximum de vraisemblance (ML) s'écrit alors

$$\hat{\ell}(y) = \arg \max_{\ell \in \{1, \dots, M\}} p_{Y|X}(y, x_\ell) \quad (9)$$

Nous montrons ci-après, que ce détecteur peut être remplacé par un détecteur plus simple. Les détecteurs déjà obtenus dans la littérature pour les cas asymptotiques (I faible et I grand) peuvent se retrouver facilement à partir du détecteur proposé. Nous commençons par un résultat intermédiaire.

Résultat 1. Soit $f_I(z)$ une fonction de \mathbb{R}_+ vers \mathbb{R} et telle que

$$f_I(z) = z^2 - \ln I_0(2\sqrt{I}z) \quad (10)$$

Soit $\hat{z}_I = \arg \min_{z \in \mathbb{R}_+} f_I(z)$, $B_{inf} = \sqrt{\max(I-1; 0)}$ et $B_{sup} = \sqrt{\max(I-1/2; 0)}$ alors, $\forall I \in \mathbb{R}_+$, $f_I(z)$ est décroissante pour $z \leq B_{inf}$, croissante pour $z \geq B_{sup}$ et $\hat{z}_I \in [B_{inf}, B_{sup}]$.

La démonstration est en annexe. Nous pouvons conclure que le récepteur optimal de l'équation (9) peut être remplacé par

$$\hat{\ell}(y) = \arg \min_{\ell \in \{1, \dots, M\}} |y - \sqrt{S}x_\ell - \hat{z}_I| \quad (11)$$

Nous pouvons remarquer que lorsque l'interférence est faible ($I < 1/2$) alors $\hat{z}_I = 0$ et le détecteur optimal devient

$$\hat{\ell}(y) = \arg \min_{\ell \in \{1, \dots, M\}} |y - \sqrt{S}x_\ell| \quad (12)$$

Ceci est en accord avec les résultats de [9] qui montrent que, lorsque l'interférence est faible, elle peut être considérée comme du bruit gaussien. En appliquant cela au modèle en (1), on retrouve bien le décodeur de l'Eq. (12). A l'autre extrémité, lorsque $I \gg 1$, on peut considérer que $\hat{z}_I \approx \sqrt{I}$. Intuitivement, à haut INR, on peut estimer θ_I par $\hat{\theta}_I = \arg(y - \sqrt{S}x)$. Après soustraction de l'interférence estimée on a

$$\tilde{y} = y - \sqrt{I}e^{j\hat{\theta}} \approx \sqrt{S}x + n \quad (13)$$

Le decodeur ML adapté à \tilde{y} est donc

$$\hat{\ell}(\tilde{y}) = \arg \min_{\ell \in \{1, \dots, M\}} |\tilde{y} - \sqrt{S}x_\ell| \quad (14)$$

Comme $e^{j\hat{\theta}_I} = \frac{y - \sqrt{S}x_\ell}{|y - \sqrt{S}x_\ell|}$, on peut écrire

$$|\tilde{y} - \sqrt{S}x_\ell| = \left| (y - \sqrt{S}x_\ell) \left(1 - \frac{\sqrt{I}}{|y - \sqrt{S}x_\ell|} \right) \right| \quad (15)$$

$$= |y - \sqrt{S}x_\ell| - \sqrt{I} \quad (16)$$

Cela correspond bien à (11) avec $\hat{z}_I = \sqrt{I}$.

5 Résultats numériques

Nous considérons ici une interférence I telle que $I = \sqrt{S}$. C'est une situation complexe dans laquelle l'interférence n'est ni négligeable, ni prépondérante. Pour entraîner le réseau, le

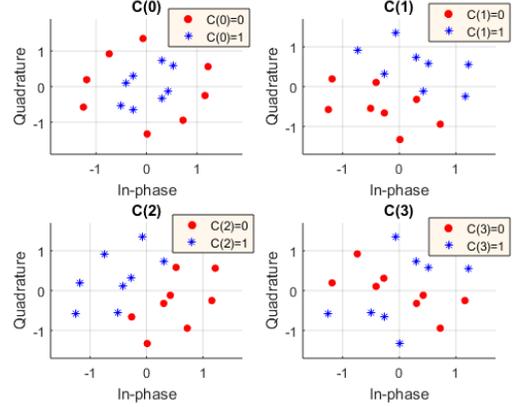


FIGURE 2 – Constellation et région de décisions obtenues pour $SNR = 20dB$, $INR = 10dB$, $k = 4$.

SNR est fixé à 20dB. La longueur des mots binaires est $k = 4$ par conséquent $|\mathcal{X}| = 16$. La constellation obtenue est donnée sur la figure 2. Pour chaque bit $C(i)$, $0 \leq i \leq 3$, les points de \mathcal{X} tels que $C(i) = 0$ sont en rouge et les autres en bleu. Nous pouvons ainsi voir les régions de décision pour chacun des bits. On peut noter une séparation claire entre les deux catégories ce qui facilite la classification binaire réalisée par le détecteur.

Nous nous intéressons maintenant au taux d'erreur binaire. Soit B le message d'information, il est codé à l'aide d'un code LDPC issu du standard DVB-S.2. La longueur totale du mot de code est 64800. Les données codées sont ensuite découpées en paquets de longueur $k = 4$ associées ensuite au point correspondant de la constellation de la figure 2. En réception, deux types de détecteurs sont envisagés. Le premier est le détecteur à base de réseau de neurones (DT-NN, Eq. (5)). Il fournit des probabilités \hat{C}_i à partir desquelles nous pouvons calculer le ratio $LLR_i = \log \frac{1 - \hat{C}_i}{\hat{C}_i}$ qui sert d'entrée au décodeur LDPC. Le second détecteur est l'approximation du détecteur optimal donné dans la section 4. Les LLR sont obtenues par

$$LLR(C_i) = \log \left(\frac{\sum_{x \in \mathcal{X}: x_i=0} p_{Y|X}(y, x)}{\sum_{x \in \mathcal{X}: x_i=1} p_{Y|X}(y, x)} \right) \quad (17)$$

avec $p_{Y|X}(y, x) \propto e^{-|y - \sqrt{S}x - \sqrt{I}e^{j\arg(y-x)}|^2}$. On notera ce détecteur DT-Q-Optim. Enfin, nous comparerons les résultats avec le détecteur usuel pour un canal AWGN (DT-AWGN) pour lequel $p_{Y|X}(y, x) \propto e^{-\frac{|y - \sqrt{S}x|^2}{1+I}}$ et avec le détecteur optimal construit avec la probabilité conditionnelle donnée dans (8). Les résultats sont donnés sur la figure 3. On peut constater que la solution proposée (\mathcal{X}_{NN} + DT-Q-Optim) amène un gain cumulé d'environ 2dB par rapport à la solution standard (DT-AWGN + 16QAM). Le détecteur optimal permet un gain de 0.7dB et la constellation un gain de 1.3dB supplémentaires. On constate enfin que le réseau de neurones fournit une approximation correcte (à 0.2dB) du détecteur optimal et que l'approximation (11) n'amène aucune dégradation par rapport au récepteur optimal. Des résultats plus complets sont donnés dans [7] confirmant la validité de l'approche proposée pour tout I .

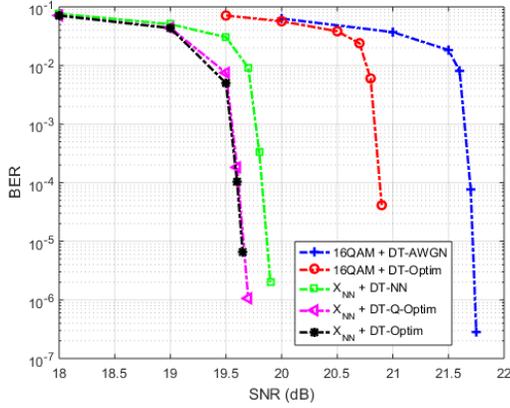


FIGURE 3 – BER vs SNR, $I = \sqrt{S}$, code LDPC de ratio $R = 5/6$, X_{NN} est la constellation de la figure 2.

6 Conclusion

Dans cet article, la coexistence entre un signal radar et un signal dédié à un autre type de communication a été étudiée. Par le biais des techniques d'apprentissage par réseau de neurones, nous avons pu optimiser des constellations pour un niveau d'interférence donné. Nous avons également proposé un détecteur (quasi)-optimal à faible coût. On a pu ainsi constater que communiquer demeure possible même en présence d'une interférence non-négligeable. Ces résultats doivent être maintenant étendus à des modèles de communication plus complexes.

Annexe

Pour montrer le résultat 1, on s'intéresse à la dérivée de la fonction $f_I(z)$. Elle a pour expression

$$f'_I(z) = 2z - 2\sqrt{I} \frac{I'_0(2\sqrt{I}z)}{I_0(2\sqrt{I}z)}$$

La dérivée de I_0 s'écrit

$$I'_0(u) = \sum_{k=0}^{+\infty} \frac{1}{k!(k-1)!} \left(\frac{u}{2}\right)^{2k-1} = I_{-1}(u)$$

Lorsque m est un entier, $I_m(z) = \frac{1}{\pi} \int_0^\pi e^{z \cos(\theta)} \cos(m\theta) d\theta$ et donc $I_{-m}(u) = I_m(u)$. Cela conduit à

$$f'_I(z) = 2z - 2\sqrt{I} \frac{I_1(2\sqrt{I}z)}{I_0(2\sqrt{I}z)}$$

On étudie le signe de $f'_I(z)$. On utilise pour cela les inégalités sur les fonctions de Bessel [10] :

$$\frac{u}{1 + \sqrt{1 + u^2}} \leq \frac{I_1(u)}{I_0(u)} \leq \frac{u}{\frac{1}{2} + \sqrt{\frac{1}{4} + u^2}} \quad u \in \mathbb{R}_+^* \quad (18)$$

Cela conduit à l'inégalité suivante

$$A_{\frac{1}{2}}(z) \leq f'_I(z) \leq A_1(z) \quad (19)$$

où $A_\beta(z) = 2z - \frac{4Iz}{\beta + \sqrt{\beta^2 + 4Iz^2}}$ et $\beta > 0$. Soit $a_\beta(z) = \beta + \sqrt{4Iz^2 + \beta^2} - 2I$, on a $\text{signe}(A_\beta(z)) = \text{signe}(a_\beta(z))$ puisque $z \in \mathbb{R}^+$. La fonction $a_\beta(z)$ est strictement croissante sur \mathbb{R}^+ par conséquent, si $I \geq \beta$, $a_\beta(z) = 0$ a une solution unique $z_0 = \sqrt{I - \beta}$ sur \mathbb{R}^+ . On en déduit $f'_I(z) \leq 0$ pour $0 \leq z \leq \sqrt{I - 1}$ et $f'_I(z) \geq 0$ pour $z \geq \sqrt{I - \frac{1}{2}}$. La fonction $f_I(z)$ atteint donc son minimum quelque part dans l'intervalle $[\sqrt{I - 1}; \sqrt{I - \frac{1}{2}}]$ et cela s'applique lorsque $I > 1$. Lorsque $I < 1/2$, on peut conclure que $f_I(z)$ est minimum en 0. Enfin, dans le cas intermédiaire, $f_I(z)$ atteint son minimum quelque part dans l'intervalle $[0; \sqrt{I - 1/2}]$. En rassemblant ces trois résultats, on obtient $\sqrt{\max(I - 1; 0)} \leq \hat{z} \leq \sqrt{\max(I - 1/2; 0)}$.

Références

- [1] Z. Khan, J. J. Lehtomaki, R. Vuottoniemi, E. Hossain, and L. A. Dasilva. On opportunistic spectrum access in radar bands : Lessons learned from measurement of weather radar signals. *IEEE Wireless Commun.*, 23(3) :40–48, June 2016.
- [2] N. Nartasilpa, A. Salim, D. Tuninetti, and N. Devroye. Communications System Performance and Design in the Presence of Radar Interference. *IEEE Trans. on Commun.*, 66(9) :4170–4185, Sept 2018.
- [3] N. Nartasilpa, D. Tuninetti, and N. Devroye. Signal constellation design in the presence of radar interference and Gaussian noise. In *IEEE Military Communications Conference (MILCOM)*, pages 719–724, Oct 2017.
- [4] N. Nartasilpa, D. Tuninetti, N. Devroye, and D. Erricolo. Let's share CommRad : Effect of radar interference on an uncoded data communication system. In *2016 IEEE Radar Conference (RadarConf)*, May 2016.
- [5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [6] T. Gruber, S. Cammerer, J. Hoydis, and S. ten Brink. On deep learning-based channel decoding. In *CISS'17*, 2017.
- [7] F. Alberge. Deep learning constellation design for the awgn channel with additive radar interference. *IEEE Trans. on Commun.*, 67(2) :1413–1423, Feb 2019.
- [8] F. Chollet. *Deep Learning with Python*. Manning, 2018.
- [9] N. Nartasilpa, D. Tuninetti, N. Devroye, and D. Erricolo. On the Error Rate of a Communication System Suffering from Additive Radar Interference. In *2016 IEEE Global Communications Conference*, pages 1–6, Dec 2016.
- [10] D. E. Amos. Computation of modified bessel functions and their ratios. *Mathematics of Computation*, 28(125) :239–251, 1974.