

Codage de Source Stratégique avec Information Adjacente au Décodeur

Maël LE TREUST¹, Samson LASAULCE¹

¹Laboratoire Signaux et Systèmes (CNRS – Université Paris Sud 11 – Supélec)

91191, Gif-sur-Yvette

{mael.letreust, samson.lasaulce}@lss.supelec.fr

Résumé – Nous caractérisons le débit d’une source d’information inspirée par la théorie des jeux. La distribution de la source dépend d’un état qui varie de manière arbitraire symbole après symbole et le décodeur observe une version dégradée des symboles réalisés par la source d’information. Ce modèle nous permet de quantifier l’information supplémentaire que nécessite un joueur dans un jeu de long-terme avec observation imparfaite des actions passées. Nous appliquons notre résultat au jeu couramment appelé ”le dilemme du prisonnier” pour lequel la région des paiements atteignable se trouve réduite.

Abstract – We characterize the rate of an information theoretic source inspired from game theory. The distribution of the source depends on a state which vary in an arbitrary manner. Moreover the decoder has a degraded version of the sequence of symbols. This model is relevant to quantify the additionnal information needed by a player during a long-run game with imperfect observation of past actions. We provide an applicative example considering ”the prisoner’s dilemma” for which the achievable payoff region is decreased.

1 Introduction

Une source d’information aux variations arbitraires dépend d’un état qui varie indépendamment de toute loi de probabilité, symbole après symbole. Nous caractérisons le débit d’une telle source, en supposant que le décodeur observe une version dégradée des symboles réalisés par la source d’information. Ce modèle se révèle pertinent lorsque l’on considère un jeu de long-terme avec observation imparfaite des actions passées [2].

A chaque étape, les joueurs $J_i \in K$ choisissent une probabilité P_i sur leurs actions X_i et l’on note $x = (x_1, \dots, x_K)$ le vecteur des actions réalisées par les joueurs, $x^n = (x_1^n, \dots, x_K^n)$ la suite de vecteurs de longueur n . Le joueur J_1 (voir figure 1) observe, à travers le canal T , une image partielle y du vecteur d’actions x réalisé. La source d’information que l’on considère est le vecteur x des actions des joueurs et l’information adjacente y représente l’observation partielle dont dispose le joueur J_1 . Un encodeur E est chargé de quantifier l’information $m \in M$ supplémentaire, nécessaire au joueur J_1 afin qu’il observe parfaitement les suites d’actions des autres joueurs.

Par hypothèse, chaque joueur joue une suite de probabilités i.i.d. et le retard d’observation dû au codage ne pénalise pas les joueurs. Le codage de source classique n’est pas adapté puisque un joueur malveillant $J_j \in K$ peut modifier la distribution de probabilité sur ses actions $P_j \in \Delta(X_j)$. Ce joueur peut choisir une suite d’actions qui induit une erreur dans le schéma de codage et modifie le résultat du jeu.

La compression d’une source où l’information est une suite d’actions choisies par des joueurs pouvant dévier de manière arbitraire n’est pas étudiée à ce jour. Les résultats d’Ahlsvede

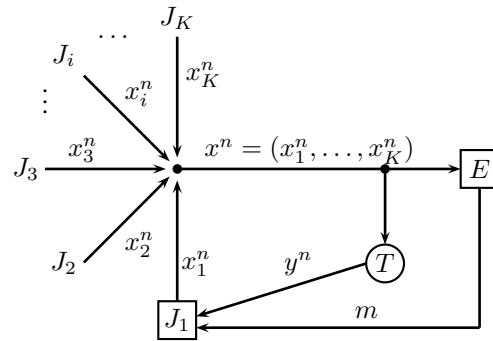


FIGURE 1 – Le joueur J_1 observe une image y des actions x .

[1] sur les sources corrélées aux variations arbitraires sont proches de nos travaux, notamment le théorème 4’ [1]. La différence principale avec nos travaux est la définition de l’état de la source et la technique de la preuve. Le fait de considérer les déviations unilatérales des joueurs nous obligent à transmettre sans erreur la suite d’actions d’un des joueurs. Notre preuve utilise une suite de coloriage de graphe où les sommets sont des symboles alors qu’Ahlsvede [1] utilise un coloriage de graphe où les sommets sont des suites de symboles. Afin d’utiliser la loi des grands nombres pour les suites de symboles, cette technique de preuve nécessite une condition de “positivité entropique” énoncée page 221 de l’article [1] :

$$H(X(s)|Y(s)) \times H(Y(s)|X(s)) > 0, \quad \forall s \in S \quad (1)$$

Dans le cas de notre source d’information, cette condition est remplacée par une condition sur les suites d’états admissibles (voir la condition 2) et par le fait que la variable aléatoire y ne

dépende de s qu'à travers x . Notre résultat s'applique aux cas des transitions T déterministes alors qu'une telle transition ne satisfait pas la contrainte d'Ahlswede.

Nous présentons, dans la section 2, le modèle de source étudié, accompagné, dans la section 3, du résultat principal qu'est le théorème 1 ainsi que de sa preuve, section 4. Ensuite, nous présenterons une application de notre résultat au cas du jeu du dilemme du prisonnier 5. Nous verrons que la région des paiements atteignables se trouve modifiée si l'on ajoute une contrainte de capacité $\mathcal{R} \leq C$ sur le débit de la source.

Par souci de clarté, on note $\Delta(X_i)$ l'ensemble des distributions de probabilité sur l'ensemble X_i . Considérons une famille d'ensembles $(X_i)_{i \in K}$ et une famille de distributions $P_i \in \Delta(X_i)$ pour chaque composante $i \in K$. Notons $X = \prod_{i \in K} X_i$ le produit cartésien; P le produit des probabilités défini par $P = \otimes_{i \in K} P_i \in \prod_{i \in K} \Delta(X_i)$ et P_{-i} la probabilité définie par $P_{-i} = \otimes_{j \neq i} P_j \in \prod_{j \neq i} \Delta(X_j)$ où la composante $i \in K$ a été retirée.

2 Modèle

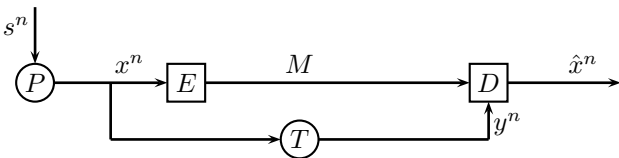
Nous étudions une source d'information aux variations arbitraires (AVS). L'information est modélisée par une variable aléatoire X dont la distribution de probabilité $P(x|s)$ dépend d'un paramètre d'état $s \in S$ qui varie de manière arbitraire symbole après symbole. La source d'information X que nous considérons est un vecteur $X = (X_1, \dots, X_K)$ de variables aléatoires. une suite d'états de la source $s^n \in S^n$ est une suite de distributions sur l'une des composantes $j \in K$ du vecteur $x^n = (x_1, \dots, x_K)^n$. On note S^n l'ensemble des suites d'états de la source.

$$S^n = \cup_{i \in K} \Delta(X_i)^{\otimes n} \quad (2)$$

Considérons par exemple, la suite $s^n = (Q_j(1), \dots, Q_j(n)) \in S^n$ qui seulement modifie la distribution de probabilité de la composante $j \in K$ du vecteur $x^n = (x_1, \dots, x_K)^n$. A l'étape $1 \leq t \leq n$ la composante $x_j(t)$ est tirée suivant $Q_j(t) \in \Delta(X_j)$ alors que les autres composantes $x_k(t)$ sont tirées suivant une distribution $P_k \in \Delta(X_k)$ iid. Le vecteur de d'information $x(t)$ est tiré, à l'étape t , avec la probabilité suivante :

$$P(x_1, \dots, x_j, \dots, x_K | s) = [P_1 \otimes \dots \otimes Q_j(t) \otimes \dots \otimes P_K](x)$$

Nous considérons la situation où le décodeur possède une information adjacente. Il observe, à travers une probabilité de transition $T : X \rightarrow \Delta(Y)$, une version dégradée y des symboles x choisis par la source d'information.



On note $s_i^n \in \Delta(X_i)^n$ une suite d'états de la source qui modifie la i -ème coordonnée du vecteur $x = (x_1, \dots, x_K)$.

Définition 1 Un (n, M) -code est défini par un couple de fonctions :

$$f : X^n \rightarrow M, \quad g : M \times Y^n \rightarrow X^n \quad (3)$$

La probabilité d'erreur \mathcal{P}_e^n associée au (n, M) -code est définie comme suit :

$$\mathcal{P}_e^n = \max_{i \in K} \max_{s_i^n \in \Delta(X_i)^n} \mathcal{P}((X^n, Y^n) \neq (\hat{X}^n, \hat{Y}^n) | s_i^n) \quad (4)$$

La probabilité d'erreur, doit être bornée supérieurement pour chaque suite $s^n \in S^n = \cup_{i \in K} \Delta(X_i)^{\otimes n}$ d'états de la source.

Définition 2 Un débit de communication R est atteignable si pour tout $\varepsilon > 0$, il existe un (n, M) -code tel que :

$$\frac{\log M}{n} \leq R + \varepsilon, \quad \mathcal{P}_e^n \leq \varepsilon \quad (5)$$

Le débit de la source \mathcal{R} est l'infimum des débits de communications atteignables.

3 Résultats principaux

Nous caractérisons le débit \mathcal{R} d'une source aux variations arbitraires avec information adjacente au décodeur. La variation arbitraire de la source nous empêche d'utiliser un codage à la Slepian et Wolf [4]. Pour quantifier la qualité de l'information adjacente, nous introduisons le concept de coloriage d'un graphe auxiliaire. Notons, x_{-i} le vecteur x privé de sa i -ème composante, P_{-i} la probabilité produit $P_{-i} = \otimes_{j \neq i} P_j$ et $Supp P_{-i}$ le support de la probabilité P_{-i} .

Définition 3 Pour chaque composante $i \in K$, le graphe auxiliaire $G_i = (X_i, E_i)$ est défini comme suit. Les symboles $x_i \in X_i$ de la composante $i \in K$ de la source sont les sommets du graphe. Il existe une arête $e_i = (x_i, x'_i) \in E_i$ entre deux sommets x_i et x'_i si et seulement si :

$$\exists x_{-i} \in Supp P_{-i}, \exists y \in Y, \exists \delta > 0, s.t. \\ \min(T(y|x_i, x_{-i}), T(y|x'_i, x_{-i})) \geq \delta$$

Définition 4 Soit C un ensemble de couleurs. Un coloriage du graphe G_i est une fonction $\phi_i : X_i \rightarrow C$ qui satisfait :

$$\forall e_i = (x_i, x'_i) \in E_i, \text{ on a } \phi_i(x_i) \neq \phi_i(x'_i)$$

Un coloriage minimal du graphe G_i est un coloriage ϕ_i pour lequel la cardinalité $|C|$ est minimale.

Définition 5 Le nombre chromatique χ_i du graphe G_i est la cardinalité de C pour le coloriage minimal du graphe G_i .

Remarque : l'information adjacente $y \in Y$ ainsi que la couleur $c \in C$ du sommet $x_i \in X_i$ dans le graphe G_i permet au décodeur de retrouver exactement le symbole $x_i \in X_i$. La propriété principale est que ce décodage ne dépend pas de la distribution de la source.

Afin d'encoder les symboles $x_{-i} \in X_{-i}$, nous définissons, pour chaque symbole $x_i \in X_i$, la probabilité de transition marginale T_{x_i} associée au symbole x_i .

$$T_{x_i} : X_{-i} \longrightarrow \Delta(Y) \quad (6)$$

$$x_{-i} \longrightarrow T_{x_i}(y|x_{-i}) = T(y|x_i, x_{-i}) \quad (7)$$

On notera Y_{x_i} le signal aléatoire, distribué suivant la loi de probabilité induite par T_{x_i} .

Théorème 1 Soit X une source d'information aux variations arbitraires et T une probabilité de transition. Le débit de la source \mathcal{R} est donné par l'équation suivante :

$$\mathcal{R} = \max_{i \in K} [\max_{x_i \in X_i} H(X_{-i}|Y_{x_i}) + \log \chi_i] \quad (8)$$

H représente l'entropie de Shannon [3] et χ_i le nombre chromatique du graphe G_i .

4 Preuve du Théorème 1

4.1 Atteignabilité

Nous construisons un codage basé sur le coloriage des graphes et les tests statistiques.

La fonction d'encodage f . L'encodeur procède à un test statistique en calculant l'ensemble suivant :

$$\arg \min_{i \in K} \sum_{x_{-i} \in X_{-i}} \left| \frac{N(x_{-i}^n | x_i)}{n} - P_{-i}(x_{-i}) \right| \quad (9)$$

et choisit une composante $i \in K$. Les symboles de la composante $i \in K$ vont être encodés en utilisant le coloriage du graphe correspondant.

- Encoder l'index de la composante choisie i en utilisant $|K|$ messages.
- Encoder la suite des couleurs c_i^n avec à chaque étape $c_i = \phi_i(x_i)$ en utilisant un encodage avec χ_i^n messages.

Les autres composantes x_{-i}^n seront encodées en fonction de la probabilité de transition T et de la suite x_i^n . Si le symbole x_i a été souvent utilisé, la suite y_{x_i} , tirée selon la probabilité T_{x_i} , est suffisamment longue pour utiliser un codage de source avec information adjacente à la Slepian and Wolf [4]. Dans le cas contraire, l'information x_{-i} doit être encodée sans aucune compression. L'encodeur divise la suite y^n en une partition $(y_{x_i})_{x_i \in X_i}$ indexée par les symboles $x_i \in X_i$ correspondant à la transition T_{x_i} . Notons $n_{x_i} = N(x_i^n | x_i)$ et $y_{x_i}^{n_{x_i}}$ la sous-suite d'informations adjacentes tirées avec T_{x_i} . L'encodeur calcule la partition $(\tilde{X}_i, \tilde{X}_i^c)$ des symboles l'ensemble $x_i \in X_i$.

- Si $N(x_i^n | x_i) = n_{x_i} \leq \bar{n}_1$, alors $x_i \in \tilde{X}_i$ et la suite x_{-i} est encodée avec $|X_{-i}|^n$ messages.
- Si $N(x_i^n | x_i) = n_{x_i} > \bar{n}_1$, alors $x_i \in \tilde{X}_i^c$ et chaque suite x_{-i} est encodée en utilisant le "binning" aléatoire avec $M_{x_i} = 2^{n_{x_i}(H(x_{-i}|y_{x_i})+2\varepsilon)}$ messages.

L'ensemble des messages \mathcal{M} d'un tel schéma de codage est de cardinalité :

$$M = |K| \cdot \chi_i^n \cdot |X_{-i}|^{\sum_{x_i \in \tilde{X}_i} n_{x_i}} \cdot \prod_{x_i \in \tilde{X}_i^c} 2^{n_{x_i}(H(x_{-i}|y_{x_i})+2\varepsilon)}$$

La fonction de décodage g . Le décodeur décode séparément les information concernant la composante i des autres composantes $-i$.

- Connaissant la composante $i \in K$ choisie lors des tests statistiques, l'information adjacente $y \in Y$ et la couleur $c_i \in \chi_i$ caractérise un unique symbole d'étape $x_i \in X_i$ de la composante i .

Le décodeur caractérise la partition \tilde{X}_i et \tilde{X}_i^c de l'ensemble des symboles X_i .

- Pour la transition T_{x_i} avec le symbole $x_i \in \tilde{X}_i$ la suite de symboles x_{-i} est décodée directement.
- Pour la transition T_{x_i} avec le symbole $x_i \in \tilde{X}_i^c$, le décodeur cherche dans le bin $m \in \mathcal{M}_{x_i}$ une suite x_{-i}^n qui soit conjointement typique avec l'information adjacente y_{x_i} pour la distribution $P_{-i} \otimes T_{x_i} \in \Delta(X_{-i} \times Y)$.

Le débit du code. Soit $\bar{n}_2 > \frac{\log |K| + \bar{n}_1 |X_i| \log |X_{-i}|}{\varepsilon}$. Nous montrons que pour tout $n \geq \bar{n}_2$, le débit du code est borné par la quantité suivante :

$$\frac{\log M}{n} \leq \max_{i \in K} [\max_{x_i \in X_i} H(x_{-i}|y_{x_i}) + \log \chi_i] + 3\varepsilon = H + 3\varepsilon$$

La probabilité d'erreur. Supposons que la distribution arbitraire ne concerne que la composante $i \in K$ du vecteur $x = (x_1, \dots, x_K)$. Lorsque l'encodeur choisit la composante $i \in K$ grace aux tests statistiques, alors la suite des vecteurs x^n est reconstruite parfaitement avec une grande probabilité.

Supposons que l'encodeur choisisse, lors des tests statistiques, une autre composante $j \neq i \in K$. Cela implique l'inégalité suivante :

$$\sum_{x_{-j} \in X_{-j}} \left| \frac{N(x_{-j} | x_j)}{n} - P_{-j}(x_{-j}) \right| \quad (10)$$

$$\leq \sum_{x_{-i} \in X_{-i}} \left| \frac{N(x_{-i} | x_i)}{n} - P_{-i}(x_{-i}) \right| \quad (11)$$

La suite x_j^n est tirés iid avec la distribution P_j , donc elle est typique avec une grande probabilité. Par conséquent, la suite de la composante x_i^n est typique avec grande probabilité et donc correctement encodée. On peut donc trouver un n suffisamment grand tel que la probabilité d'erreur soit bornée par ε .

4.2 Réciproque

Pour montrer la réciproque, nous supposons que le débit R soit inférieur à la quantité suivante.

$$R < \max_{i \in K} [\max_{x_i \in X_i} H(X_{-i}|Y_{x_i}) + \log \chi_i] \quad (12)$$

Supposons que le maximum soit atteint pour la composante $i \in K$ et le symbole $\hat{x}_i \in X_i$. La distribution sur le vecteur d'information $x = (x_1, \dots, x_K)$ est le produit des distributions marginales sur chaque composante $x_i, i \in K$. Le débit optimal R s'écrit donc comme la somme des débits des composantes $R = \sum_{i \in K} R_i$. La réciproque est divisée en deux parties.

Tout d'abord, supposons que le schéma de codage ait un débit $R_i < \log \chi_i$. Par la propriété du coloriage minimal, la probabilité d'erreur tend vers 1 lorsque n tend vers l'infini.

$$\exists \tilde{x}_i, \tilde{x}'_i, \exists x_{-i} \in \text{Supp } P_{-i}, \exists y \in Y, \exists \delta > 0, \\ \min(T(y|x_i, x_{-i}), T(y|x'_i, x_{-i})) \geq \delta$$

$$\implies \forall (f, g), \mathcal{P}_e^n > \sum_{t=1}^n \delta^2 (1 - \delta^2)^{t-1} \xrightarrow{n \rightarrow +\infty} 1$$

Il existe une suite d'états de la source telle que les symboles \tilde{x}_i et \tilde{x}'_i ne puissent être distingués par le décodeur. Pour une telle suite, la procédure de codage induit une erreur qui tend vers 1.

Ensuite, supposons que la suite x_i soit parfaitement reconstruite au décodeur mais que le débit de la suite restante soit :

$$R_{-i} < H(x_{-i}|y_{x_i}) = \max_{x_i \in X_i} H(X_{-i}|Y_{x_i}) \quad (13)$$

Dans ce cas, la probabilité d'erreur tend vers 1 lorsque la suite d'états $s_i^n = \tilde{x}_i^n$ et que n tend vers l'infini.

Ceci contredit l'hypothèse que le débit suivant soit atteignable.

$$R = R_{-i} + R_i < \max_{x_i \in X_i} H(x_{-i}|y_{x_i}) + \log \chi_i \quad (14)$$

5 Une application à la théorie des jeux : le dilemme du prisonnier

Nous considérons la situation décrite figure 1 à deux joueurs, où un encodeur veut quantifier l'information x que le joueur 1 pourra correctement décodeur. Le jeu est défini par la matrice de paiements décrite Fig. (2) et la probabilité de transition est définie Fig. (3). Par exemple, la probabilité de recevoir le signal “ s_{hg} ” est de $1 - \frac{3}{4}\varepsilon$, lorsque l'action (h, g) est jouée. Cette transition est définie de manière symétrique pour les autres couples d'actions (h, d) , (b, g) et (b, d) .

	g	d
h	3, 3	0, 4
b	4, 0	1, 1

FIGURE 2 – La matrice des paiements pour le “dilemme du prisonnier”.

Les joueurs choisissent leur action avec une probabilité $P_1 \in \Delta(\{h, b\})$ et $P_2 \in \Delta(\{g, d\})$. Si l'on considère le codage de source avec information adjacente de Slepian et Wolf [4], un joueur peut avoir intérêt à changer la distribution de probabilité sur ses actions de manière à manipuler le schéma du codage. Nous garantissons un schéma codage fiable et robuste à toute déviation unilatérale des joueurs.

En guise d'illustration, nous fixons le paramètre $\varepsilon = 0.5$ de la probabilité de transition Fig. (3) et nous imposons une contrainte de capacité sur le débit de la source $\mathcal{R} \leq C = 1,7$

$$T(h, g) = \begin{array}{|c|c|} \hline 1 - \frac{3}{4}\varepsilon & \frac{1}{4}\varepsilon \\ \hline \frac{1}{4}\varepsilon & \frac{1}{4}\varepsilon \\ \hline \end{array} \begin{array}{l} s_{hg} \quad s_{hd} \\ s_{bg} \quad s_{bd} \end{array}$$

FIGURE 3 – La probabilité de transition lorsque l'action (h, g) est jouée.

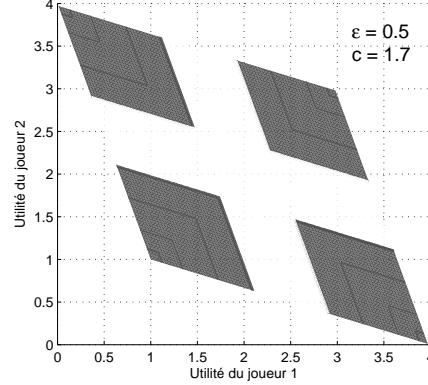


FIGURE 4 – Région des utilités atteignables pour le dilemme du prisonnier avec une contrainte de capacité.

bits par seconde. La figure (3) représente la région des paiements espérés des joueurs pour lesquels la distribution de probabilité satisfait la contrainte de capacité.

6 Conclusion

Nous étudions une classe de sources aux variations arbitraires avec information adjacente au décodeur. Nous caractérisons le débit optimal d'une telle source et montrons que ce résultat s'applique au cas d'un jeu dans lequel les joueurs peuvent dévier de manière unilatérale. Notre résultat garantit qu'aucun joueur ne peut manipuler le schéma du codage en changeant la distribution de probabilité sur ses actions.

Références

- [1] R. Ahlswede. Coloring hypergraphs : A new approach to multi-user source coding, part 2. *Journal of combinatorics, information and system sciences*, 5(3) :220–268, 1980.
- [2] J. Renault and T. Tomala. General properties of long-run supergame. *Dynamic Games and Applications*, 1(2) :319–350, 2011.
- [3] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 :379–423, Jul. 1948.
- [4] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IT-19* :471–480, July. 1973.