

L'entropie comme mesure de difficulté des problèmes inverses

Florent MICHEL¹ Benoît MALÉZIEUX¹ Matthieu KOWALSKI² Thomas MOREAU¹

¹Université Paris-Saclay, Inria, CEA
91120, Palaiseau, France

²Université Paris-Saclay, CNRS, LISN
91190, Gif-sur-Yvette, France

Résumé – Dans le contexte du problème inverse, nous proposons une limite inférieure estimable sur la façon dont nous pouvons récupérer le signal d'intérêt avec l'accès à certains échantillons. Cette limite inférieure est basée sur l'entropie conditionnelle et nous proposons une méthode pour la quantifier en haute dimension avec une approche par sections. Nous montrons que cette mesure est corrélée avec la difficulté d'un problème et qu'elle peut être convertie avec une perte de dimension équivalente.

Abstract – In the context of inverse problem, we propose a tractable lower bound on how well we can recover the signal of interest with access to some samples. This lower bound is based on conditional entropy and we propose a method to quantify it in high dimensional settings with a per-slice approach. We show that this measure correlates with the difficulty of a problem and that it can be converted as an equivalent dimension loss.

1 Introduction

Les problèmes inverses consistent à retrouver un signal inconnu à partir d'observations possiblement bruitées. Leur résolution est un enjeu majeur dans une multitude de domaines, de l'astrophysique à l'imagerie médicale. Plus spécifiquement, nous considérons le cas où les signaux observés $\mathbf{Y} \in \mathbb{R}^m$ sont la somme d'une transformation $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ de signaux d'intérêt $\mathbf{X} \in \mathbb{R}^n$ et d'un bruit blanc $\mathbf{B} \in \mathbb{R}^m$

$$\mathbf{Y} = A(\mathbf{X}) + \mathbf{B} . \quad (1)$$

On cherche à reconstruire \mathbf{X} à partir de \mathbf{Y} et de A qui est supposé connu. Dans le cas supervisé, des algorithmes d'apprentissage permettent de résoudre ces problèmes après entraînement sur une base de signaux non dégradés [11]. Les performances de ces méthodes sont fortement dépendantes de la réduction de dimension due à l'opérateur de mesure A , de son conditionnement, ainsi que de la distribution des données que l'on cherche à reconstruire. Cependant, les métriques existantes sur la qualité de la reconstruction ne donnent pas d'information sur l'optimalité de l'algorithme de reconstruction ou sur la difficulté du problème.

Dans d'autres domaines comme la classification [2] ou l'optimisation [9], il est courant de calculer la performance optimale d'un algorithme sous la forme d'un taux d'erreur de Bayes ou d'une borne inférieure sur sa vitesse de convergence. Sur ce modèle, on souhaite quantifier la difficulté d'un problème inverse en fonction de l'opérateur de mesure A et de la distribution des signaux X . Cette mesure doit répondre à plusieurs critères pour être fonctionnelle en pratique. Premièrement, elle doit être calculable facilement, y compris en grande dimension, sans avoir besoin de résoudre le problème inverse associé. Deuxièmement, elle ne doit dépendre que des données disponibles, et pas d'information a priori sur leur distribution.

Récemment, des travaux ont étudié la difficulté d'estimation de paramètres dans des systèmes physiques avec des outils issus de la théorie de l'information. Ainsi, l'information mutuelle permet de quantifier la perte d'information au cours du

temps dans un système chaotique [8], ou encore de caractériser l'identifiabilité des paramètres d'un système dynamique non linéaire en petite dimension [10]. Ces approches prometteuses, bien que construites pour des problèmes à faible nombre de paramètres, offrent des perspectives nouvelles à l'étude des problèmes inverses, notamment en grande dimension.

Contributions. Dans un cadre supervisé où l'on a accès à un jeu de données $(x_i)_{1 \leq i \leq N} \in \mathbb{R}^n$ i.i.d. suivant une distribution de probabilité de densité $p_{\mathbf{X}}(x)$ et un opérateur de mesure A permettant de générer les observations $(y_i)_{1 \leq i \leq N} \in \mathbb{R}^m$, nous proposons une mesure de la difficulté du problème inverse associé. La méthode repose sur l'estimation de l'entropie conditionnelle intervenant dans une borne inférieure sur la performance optimale d'un algorithme de reconstruction. Afin de passer à l'échelle, nous proposons une extension aux problèmes en grande dimension à l'aide de l'entropie par sections et validons l'approche sur des données simulés.

2 L'entropie comme mesure de la difficulté d'un problème inverse

Dans cette section, nous présentons une méthode pour quantifier la difficulté d'un problème inverse étant donné une distribution de signaux et un opérateur de mesure, construite sur l'entropie. On se place dans un cadre supervisé où l'on a accès à un jeu de données $(x_i)_{1 \leq i \leq N} \in \mathbb{R}^n$ i.i.d. suivant une distribution de probabilité de densité $p_{\mathbf{X}}(x)$.

Dans toute la suite, sauf mention contraire, on supposera que les variables aléatoires admettent une densité par rapport à la mesure de Lebesgue et on fixe \mathbf{X} (resp. \mathbf{Y} une telle variable sur \mathbb{R}^n (resp. \mathbb{R}^m) de densité $p_{\mathbf{X}}(x)$ (resp. $p_{\mathbf{Y}}(y)$). De plus, on supposera que toutes les quantités sont finies. On rappelle la définition de l'entropie différentielle [3] d'une variable aléatoire continue.

Definition 2.1 (Entropie) L'entropie de \mathbf{X} est définie comme suit

$$\mathcal{H}(\mathbf{X}) = - \int_{\mathbb{R}^n} p_{\mathbf{X}}(x) \log(p_{\mathbf{X}}(x)) dx \quad (2)$$

Cette quantité caractérise l'incertitude d'une distribution de probabilité. Plus l'entropie est faible, plus la variable est confinée dans un petit espace avec grande probabilité. De manière similaire, on peut définir une mesure de l'incertitude d'une distribution a posteriori à l'aide de l'entropie conditionnelle.

Definition 2.2 (Entropie conditionnelle) L'entropie de \mathbf{X} conditionnellement à \mathbf{Y} est définie comme suit

$$\mathcal{H}(\mathbf{X}|\mathbf{Y}) = \int_{\mathbb{R}^m} \mathcal{H}(\mathbf{X}|\mathbf{Y} = y) p_{\mathbf{Y}}(y) dy \quad (3)$$

où $\mathcal{H}(\mathbf{X}|\mathbf{Y} = y)$ est l'entropie de la distribution de \mathbf{X} conditionnellement à $\mathbf{Y} = y$, dont la densité est $p_{\mathbf{X}|\mathbf{Y}=y}(x) = p_{\mathbf{X}|\mathbf{Y}}(x|y)$.

L'entropie conditionnelle permet de quantifier l'incertitude moyenne sur la variable \mathbf{X} sachant que l'on connaît la valeur de \mathbf{Y} . Dans le contexte des problèmes inverses, cette quantité est particulièrement intéressante pour quantifier l'incertitude sur les données \mathbf{X} sachant les observations \mathbf{Y} . En exploitant la règle de Bayes, on obtient une expression permettant le calcul de l'entropie conditionnelle de \mathbf{X} sachant \mathbf{Y} en pratique, comme décrit dans la Proposition 2.3.

Proposition 2.3 Soit $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B}$ où $\mathbf{A} \in \mathbb{R}^{m \times n}$ et \mathbf{B} est un terme de bruit indépendant de \mathbf{X} . On a alors

$$\mathcal{H}(\mathbf{X}|\mathbf{Y}) = \mathcal{H}(\mathbf{B}) + \mathcal{H}(\mathbf{X}) - \mathcal{H}(\mathbf{Y}) \quad (4)$$

Exemple : données gaussiennes. En supposant que les données suivent une distribution gaussienne multivariée et que le bruit est gaussien, on peut expliciter l'entropie a priori et l'entropie conditionnelle comme suit.

Proposition 2.4 Soit $\mathbf{X} \sim \mathcal{N}(0, \Sigma)$ où $\Sigma \in S_n^{++}(\mathbb{R})$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ et $\mathbf{B} \sim \mathcal{N}(0, \sigma^2 I_m)$. On a alors

$$\mathcal{H}(\mathbf{X}) = \frac{1}{2} \log [(2\pi e)^n \det(\Sigma)] \quad (5)$$

$$\mathcal{H}(\mathbf{X}|\mathbf{Y}) = \mathcal{H}(\mathbf{X}) - \frac{1}{2} \log \det \left(I_m + \frac{1}{\sigma^2} \mathbf{A}\Sigma\mathbf{A}^T \right) \quad (6)$$

L'entropie conditionnelle est croissante en la variance du bruit, et dépend des valeurs propres de la matrice $\mathbf{A}\Sigma\mathbf{A}^T$.

Borne de l'erreur de reconstruction en problème inverse.

Dans le contexte des problèmes inverses, étant donné un estimateur $\hat{\mathbf{X}}(\mathbf{Y})$ des données \mathbf{X} à partir des observations \mathbf{Y} , on peut mesurer l'erreur de reconstruction grâce à l'erreur quadratique moyenne

$$\text{EQM}(\hat{\mathbf{X}}(\mathbf{Y}), \mathbf{X}) = \mathbb{E}_{(\mathbf{X}, \mathbf{Y})} [\|\mathbf{X} - \hat{\mathbf{X}}(\mathbf{Y})\|_2^2] \quad (7)$$

Cette quantité dépend de la distribution de \mathbf{X} ainsi que de l'opérateur \mathbf{A} qui relie \mathbf{X} à \mathbf{Y} . Ces derniers étant fixés, on se demande quelle est l'erreur quadratique minimale du meilleur estimateur $\hat{\mathbf{X}}(\mathbf{Y})$. On rappelle un résultat de théorie de l'information [3] qui fournit une borne inférieure sur l'erreur quadratique moyenne en fonction de l'entropie conditionnelle.

Proposition 2.5 Soit $\hat{\mathbf{X}}(\mathbf{Y})$ un estimateur de \mathbf{X} . On a alors

$$\frac{1}{n} \mathbb{E}_{(\mathbf{X}, \mathbf{Y})} [\|\mathbf{X} - \hat{\mathbf{X}}(\mathbf{Y})\|_2^2] \geq \frac{1}{2\pi e} e^{\frac{2}{n} \mathcal{H}(\mathbf{X}|\mathbf{Y})} \quad (8)$$

L'entropie conditionnelle fournit une borne inférieure sur l'erreur quadratique moyenne que le meilleur algorithme de reconstruction peut atteindre. C'est donc une mesure de la difficulté du problème : plus elle est élevée, moins l'erreur quadratique moyenne associée à $\hat{\mathbf{X}}$ peut être faible.

Echelle de difficulté en pratique. L'entropie conditionnelle en elle-même est difficilement interprétable. Afin d'interpréter sa valeur, on peut la comparer à l'entropie conditionnelle d'un autre problème inverse dont on maîtrise les propriétés. Pour $\mathbf{X}_{\mathbf{g}} \sim \mathcal{N}(0, I_n)$, on introduit

$$\mathbf{Y}_{\mathbf{g}}^{(d)} = \mathbf{A}_d \mathbf{X}_{\mathbf{g}} + \mathbf{B}^{(d)} \quad (9)$$

où $\mathbf{A}_d \in \mathbb{R}^{(n-d) \times n}$ est la matrice qui sélectionne les $n - d$ premières coordonnées et $\mathbf{B}^{(d)} \sim \mathcal{N}(0, \sigma^2 I_{n-d})$. D'après la Proposition 2.4, on a

$$\mathcal{H}(\mathbf{X}_{\mathbf{g}}|\mathbf{Y}_{\mathbf{g}}^{(d)}) = \frac{n}{2} \log(2\pi e) - \frac{n-d}{2} \log \left(1 + \frac{1}{\sigma^2} \right) \quad (10)$$

On peut alors comparer la difficulté du problème avec celle de (9) afin d'avoir la dimension effective perdue par l'opérateur. On obtient

$$d_g = \frac{1}{\log \left(1 + \frac{1}{\sigma^2} \right)} \left[2\mathcal{H}(\mathbf{X}|\mathbf{Y}) + n \log \left(\frac{1 + \sigma^2}{2\pi e \sigma^2} \right) \right] \quad (11)$$

qui satisfait $\mathcal{H}(\mathbf{X}|\mathbf{Y}) = \mathcal{H}(\mathbf{X}_{\mathbf{g}}|\mathbf{Y}_{\mathbf{g}}^{(d_g)})$. Dans l'esprit de [10], on peut également interpréter la borne de (8) comme une variance équivalente. Soit $\mathbf{X}_0 \sim \mathcal{N}(0, \sigma_0^2 I_n)$. On a alors

$$\mathcal{H}(\mathbf{X}|\mathbf{Y}) = \mathcal{H}(\mathbf{X}_0) \Leftrightarrow \sigma_0^2 = \frac{1}{2\pi e} e^{\frac{2}{n} \mathcal{H}(\mathbf{X}|\mathbf{Y})} \quad (12)$$

En moyenne, la distribution à posteriori de \mathbf{X} sachant $\mathbf{Y} = y$ est aussi dispersée qu'une gaussienne isotrope de variance σ_0^2 .

Estimation Etant donné $(x_i)_{i=1, \dots, N}$ des échantillons i.i.d suivant une distribution de densité $p_{\mathbf{X}}(x)$, il nous reste à estimer numériquement $\mathcal{H}(\mathbf{X}|\mathbf{Y})$. Grâce à (4) et (6), il suffit d'estimer $\mathcal{H}(\mathbf{X})$ et $\mathcal{H}(\mathbf{Y})$. Après génération d'observations $(y_i)_{i=1, \dots, N}$ telles que $y_i = \mathbf{A}x_i + b_i$ où $b_i \sim \mathcal{N}(0, \sigma^2 I_m)$, on peut estimer les entropies de façon non paramétrique avec les distances aux plus proches voisins [7]. La méthode consiste à approcher la densité $p_{\mathbf{X}}(x_i)$ de chaque point avec une valeur inversement proportionnelle au volume de la boule de rayon égal à la distance à son k-ème plus proche voisin parmi x_1, \dots, x_N . En notant $x_i(k)$ le k-ème plus proche voisin de x_i , on obtient l'estimateur de Monte-Carlo suivant, où $C_{N,k,n}$ est une constante

$$\hat{\mathcal{H}}_N^{(k)}(x_1, \dots, x_N) = C_{N,k,n} + \frac{n}{N} \sum_{i=1}^N \log (\|x_i - x_i(k)\|_2) \quad (13)$$

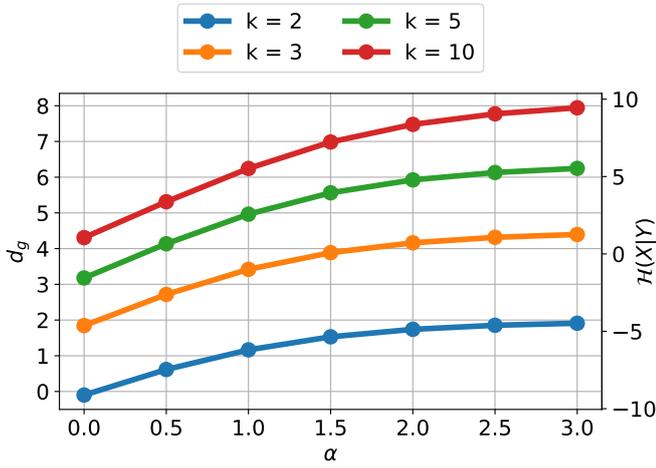


FIGURE 1 : Lorsque α augmente, le conditionnement du problème diminue, ce qui rend le problème plus dur. Aussi, la distribution des données \mathbf{X} devient plus complexe lorsque le nombre d’atomes k sélectionnés augmente. Dans cette figure, on voit que la dimension équivalente perdue d_g est croissante en α et k , conformément à l’intuition.

Résultats expérimentaux Nous démontrons expérimentalement que l’estimateur quantifie la difficulté d’un problème inverse. Des données synthétiques en dimension $n = 10$ sont générées à partir d’un dictionnaire de $L = 2n = 20$ atomes gaussiens. Pour une taille de support k donnée, un échantillon x_i est généré en sélectionnant un support aléatoire de taille k puis en tirant des coefficients gaussiens. On a donc

$$x_i = \sum_{j \in S_i} a_j^{(i)} D_j \quad (14)$$

où S_i est une partie aléatoire de $\{1, \dots, L\}$ de taille k , $D_j = X_j / \|X_j\|_2$ avec $X_j \sim \mathcal{N}(0, I_n)$ et $a_j^{(i)} \sim \mathcal{N}(0, 1)$. On étudie ensuite l’impact du conditionnement de l’opérateur sur la difficulté du problème inverse. Pour cela, on fixe $m = \lfloor n/2 \rfloor = 5$ et on tire uniformément deux matrices unitaires $U \in \mathbb{R}^{m \times m}$ et $V \in \mathbb{R}^{n \times m}$. Pour $\alpha > 0$, on considère ensuite l’opérateur

$$A_\alpha = U \text{diag}(1, 1/2^\alpha, \dots, 1/m^\alpha) V^T \quad (15)$$

dont les valeurs singulières décroissent en $1/i^\alpha$, et un bruit gaussien isotrope d’écart type $\sigma = 0.1$. Dans la Figure 1, on trace la dimension perdue équivalente d_g en fonction de k et α pour des jeux de $N = 10000$ données synthétiques. Quand la parcimonie du signal diminue (en augmentant le nombre d’atomes k sélectionnés) ou quand le conditionnement de la matrice se détériore (en augmentant α), d_g augmente car l’entropie conditionnelle (estimée) du problème augmente. Cela montre que l’entropie conditionnelle est une mesure de difficulté qui capture le conditionnement du problème ainsi que la complexité de la distribution des données.

3 Estimation en grande dimension par projection unidimensionnelle

L’estimateur d’entropie basé sur les plus proches voisins nécessite un nombre d’échantillons exponentiel en la dimension afin de converger vers la valeur de l’entropie, comme montré dans

la Figure 2. L’entropie conditionnelle définie dans la section précédente n’est donc pas adaptée pour estimer la difficulté de problèmes inverses en grande dimension. Dans cette section, nous proposons une autre mesure plus adaptée à la grande dimension basée sur des projections unidimensionnelles, et illustrons son utilité en pratique.

Entropie par sections. Quand la dimension augmente, il est possible de recourir à des projections unidimensionnelles des distributions pour mesurer des divergences [6], comme avec la distance de Sliced-Wasserstein en transport optimal, ou encore avec l’information mutuelle par sections [5] qui mesure la dépendance entre deux variables aléatoires en grande dimension. Dans ce dernier cas, les auteurs proposent de calculer la moyenne de l’information mutuelle entre des projections unidimensionnelles aléatoires des deux variables. Cette définition n’étant pas adaptée au contexte des problèmes inverses, nous proposons d’utiliser une quantité qui prend en compte le lien entre \mathbf{X} et \mathbf{Y} basée sur l’entropie par sections définie dans [5].

Définition 3.1 (Entropie par sections) Soit $\Theta \sim \mathcal{U}(\mathbb{S}^{n-1})$ un vecteur aléatoire uniformément distribué sur la sphère unité de \mathbb{R}^n . L’entropie par sections de \mathbf{X} est définie comme suit

$$\mathcal{SH}(\mathbf{X}) = \mathbb{E}_\Theta [\mathcal{H}(\Theta^T \mathbf{X})] \quad (16)$$

$$= \frac{1}{\text{Vol}(\mathbb{S}^{n-1})} \int_{\mathbb{S}^{n-1}} \mathcal{H}(\theta^T \mathbf{X}) d\theta \quad (17)$$

Pour une distribution gaussienne isotrope, \mathcal{SH} coïncide avec \mathcal{H} , ce qui n’est pas vrai en général.

Estimation. Pour estimer l’entropie par sections à partir de $(x_i)_{i=1, \dots, N}$ nous utilisons un estimateur de l’entropie en dimension 1 $\hat{\mathcal{H}}_N$ [4, 1] et appliquons une méthode de Monte-Carlo pour approcher l’espérance sur la sphère. On a alors

$$\hat{\mathcal{S}}\mathcal{H}_N(x_1, \dots, x_N) = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{\mathcal{H}}_N(\theta_i^T x_1, \dots, \theta_i^T x_N) \quad (18)$$

où les $(\theta_i)_{i=1, \dots, N_s}$ sont N_s échantillons i.i.d. de $\mathcal{U}(\mathbb{S}^{d-1})$. En pratique, contrairement à l’entropie classique, l’estimation de l’entropie par sections peut être implémentée efficacement en parallélisant les estimations d’entropie sur chaque projection.

Mesure de la difficulté en grande dimension. A partir de l’entropie par sections, nous définissons une mesure de difficulté que l’on peut estimer en grande dimension.

Définition 3.2 Soit $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B}$ où $\mathbf{A} \in \mathbb{R}^{m \times n}$ et \mathbf{B} est un terme de bruit indépendant de \mathbf{X} . On définit :

$$\mathcal{D}(\mathbf{X}|\mathbf{Y}) = n\mathcal{SH}(\mathbf{X}) - m\mathcal{SH}(\mathbf{Y}) + m\mathcal{SH}(\mathbf{B}) \quad (19)$$

Pour $\mathbf{X} \sim \mathcal{N}(0, I_n)$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ la matrice qui sélectionne les m premières coordonnées et $\mathbf{B} \sim \mathcal{N}(0, \sigma^2 I_m)$, on a $\mathcal{H}(\mathbf{X}|\mathbf{Y}) = \mathcal{D}(\mathbf{X}|\mathbf{Y})$ puisqu’alors \mathbf{X} , \mathbf{Y} et \mathbf{B} sont des gaussiennes isotropes, pour lesquelles on a $\mathcal{SH}(\mathbf{X}) = \frac{1}{n}\mathcal{H}(\mathbf{X})$ d’après la Proposition 2.4. Cette nouvelle définition s’inspire donc de l’expression de l’entropie conditionnelle d’un problème inverse donnée par (4) et les deux quantités coïncident dans le cas gaussien. Cependant, puisque $\mathcal{D}(\mathbf{X}|\mathbf{Y})$ est construite à partir de l’entropie par sections, son estimation pour des problèmes en grande dimension est plus facile.

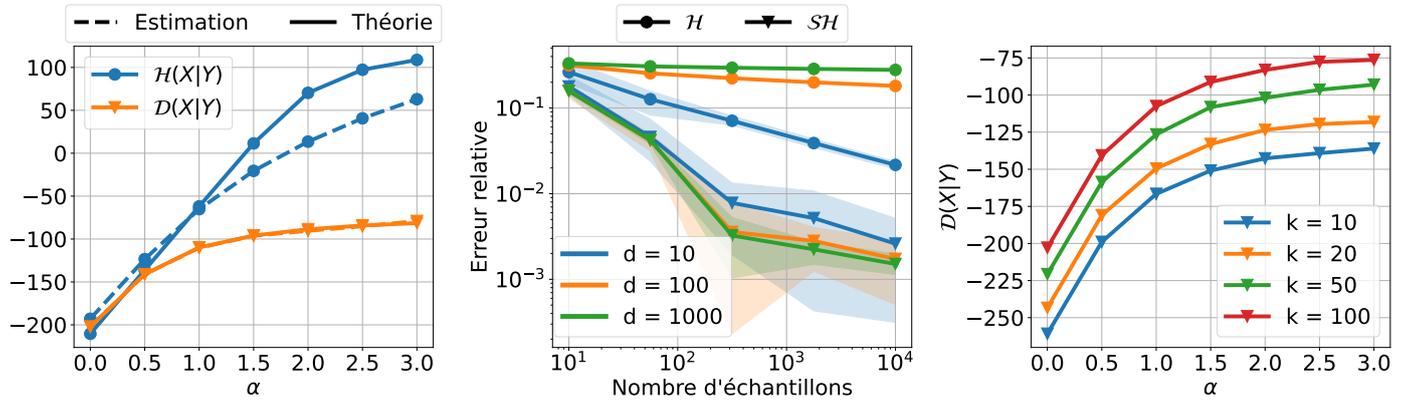


FIGURE 2 : **Gauche.** En dimension 100, pour une gaussienne non isotrope, on fait une erreur non négligeable en estimant l’entropie conditionnelle $\mathcal{H}(X|Y)$. Cependant, pour $\mathcal{D}(X|Y)$, on fait une erreur négligeable en utilisant $N_s = 1000$ projections pour estimer les entropies par sections. **Milieu.** Lorsqu’on augmente la dimension, on n’arrive plus à estimer l’entropie avec un nombre raisonnable d’échantillons à cause du fléau de la dimension. Pour l’entropie par sections, en revanche, l’estimateur converge indépendamment de la dimension des données. **Droite.** Sur des données synthétiques, $\mathcal{D}(X|Y)$ en dimension 100 a un comportement similaire à $\mathcal{H}(X|Y)$ en dimension 10 (cf. Figure 1), ce qui montre empiriquement qu’on peut capturer la difficulté du problème en ayant recours à des projections unidimensionnelles des données.

Résultats expérimentaux. Nous montrons expérimentalement que cette nouvelle mesure de difficulté est utilisable en grande dimension, tout en gardant un comportement similaire à l’entropie conditionnelle. Pour cela, on compare d’abord l’estimation de $\mathcal{H}(X|Y)$ et $\mathcal{D}(X|Y)$ pour $N = 10000$ données en dimension $n = 100$ et des observations en dimension $m = 50$ dans le cas d’une distribution gaussienne multivariée de matrice de covariance $\Sigma = \text{diag}(0.1, \dots, 2)$ où les coefficients diagonaux augmentent linéairement entre 0.1 et 2. Sur la Figure 2, on compare l’estimation de $\mathcal{H}(X|Y)$ et $\mathcal{D}(X|Y)$ et leurs valeurs théoriques pour différentes valeurs de α pour l’opérateur A_α défini en (15) et pour $\sigma = 0.001$. Les entropies sont calculées avec un estimateur par 5 plus proches voisins et les entropies par sections sont calculées avec $N_s = 1000$ sections. L’estimation des entropies par sections est beaucoup plus rapide et l’erreur sur la mesure de difficulté est plus faible que pour l’entropie classique.

Ensuite, on compare l’impact de la dimension des données sur l’estimation d’entropie et d’entropie par sections. Pour cela, on considère la distribution gaussienne multivariée définie précédemment pour plusieurs dimensions des données, puis on calcule l’erreur relative entre l’estimation et la valeur théorique en fonction du nombre d’échantillons. L’expérience est répétée 20 fois pour obtenir des quantiles à 10% et 90%. On observe que contrairement à l’entropie par sections, on n’arrive pas à estimer l’entropie en grande dimension.

Enfin, on effectue la même expérience que sur la Figure 1 en dimension $n = 100$ avec $\sigma = 0.001$ et on calcule la mesure de difficulté $\mathcal{D}(X|Y)$ en utilisant $N_s = 1000$ sections pour estimer les entropies par sections. On voit alors que dans ce cas, $\mathcal{D}(X|Y)$ se comporte de façon similaire à l’entropie conditionnelle pour un problème semblable mais en plus grande dimension.

4 Conclusion

Nous avons proposé une mesure de difficulté des problèmes inverses, avec une extension estimable en grande dimension.

Bien que les résultats empiriques illustrent l’utilité d’une telle approche, nos futurs travaux considéreront les garanties que l’on peut obtenir pour que $\mathcal{D}(X|Y)$ soit une borne inférieure sur l’erreur de reconstruction. De plus, la validation de cette approche sur des problèmes de très large échelle, notamment en imagerie, nécessitera d’adapter l’approche par projection pour prendre en compte la structure convolutive et hiérarchique des données.

Références

- [1] J-F BERCHER et Christophe VIGNAT : Estimating the entropy of a signal with applications. *IEEE transactions on signal processing*, 48(6):1687–1694, 2000.
- [2] Christopher M BISHOP et Nasser M NASRABADI : *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [3] Thomas M COVER : *Elements of information theory*. John Wiley & Sons, 1999.
- [4] Nader EBRAHIMI, Kurt PFLUGHOEFT et Ehsan S SOOFI : Two measures of sample entropy. *Statistics & Probability Letters*, 20(3):225–234, 1994.
- [5] Ziv GOLDFELD et Kristjan GREENEWALD : Sliced mutual information : A scalable measure of statistical dependence. *Advances in Neural Information Processing Systems*, 34:17567–17578, 2021.
- [6] Soheil KOLOURI, Kimia NADJABI, Umut SIMSEKLI, Roland BADEAU et Gustavo ROHDE : Generalized sliced wasserstein distances. *Advances in neural information processing systems*, 32, 2019.
- [7] Lyudmyla F KOZACHENKO et Nikolai N LEONENKO : Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, 23(2):9–16, 1987.
- [8] Kieran MURPHY et Dani BASSETT : The information bottleneck and the double pendulum : Using machine learning to study how chaos destroys information. *Bulletin of the American Physical Society*, 2023.
- [9] Arkadii Semenovich NEMIROVSKY et David Berkovich YUDIN : *Problem complexity and method efficiency in optimization*. Wiley-Interscience series in discrete mathematics. Wiley, 1983.
- [10] Sanjay PANT et Damiano LOMBARDI : An information-theoretic approach to assess practical identifiability of parametric dynamical systems. *Mathematical biosciences*, 268:66–79, 2015.
- [11] JH RICK CHANG, Chun-Liang LI, Barnabas POZOS, BVK VIJAYA KUMAR et Aswin C SANKARANARAYANAN : One network to solve them all—solving linear inverse problems using deep projection models. *In Proceedings of the IEEE International Conference on Computer Vision*, pages 5888–5897, 2017.