Rapport sur le sujet de Maths 2023 XUSR filière PSI

Le sujet cette année tournait autour du transport de masse dans sa version régularisée introduite en 2013 par Marco Cuturi dans le papier *Sinkhorn Distances : Lightspeed Computation of Optimal Transport*. La version relaxée de Kantorovitch conduit à un problème linéaire sur le simplex sous contrainte qui est difficile à résoudre en pratique pour des problèmes de grande taille. La version avec régularisation entropique préserve la convexité du problème et permet par dualité lagrangienne d'aboutir à un algorithme de résolution par minimisation alternée sur les variables duales (Sinkhorn Algorithm).

Dans la première partie on s'intéresse aux fonctions convexes sur un ensemble convexe de $\mathbf{R}^m$: on y démontre le théorème des extrema liés et on s'intéresse au point selle du Lagrangien associé au problème d'optimisation sous contraintes.

Le but de la deuxième partie est de prouver une version simplifiée du théorème du codage de la source de Shannon. On y introduit des fonctions usuelles en théorie de l'information, comme l'entropie et la divergence de Kullback-Leibler.

La troisième partie aborde le transport régularisé, thématique très en vogue en machine learning (Wasserstein GAN, transfert learning,...). On y étudie le minimum d'une fonction strictement convexe définie à partir de la divergence de Kullback-Leibler que l'on obtient, dans la partie IV, à l'aide de l'algorithme de Sinkhorn utilisant la dualité lagrangienne.

Partie I Convexité et points selles

(1) La première partie de la question a été, presque toujours, bien traitée, en revanche l'unicité n'a été correctement démontrée que par une moitié des candidats.

(2) Pas de problème particulier pour (a), en revanche dans (b) il fallait bien écrire que x était orthogonal à tout élément de l'image de A^T pour conclure. Dans (c) les candidats qui se sont lancés dans des arguments de dimension ont, pour la plupart, pris $n = m$: peu ont pensé à montrer directement l'inclusion réciproque.

(3) Les parties (b) et (c) ont été plutôt bien traitées, en revanche (a) n'a presque pas été abordée : bien que ce soit très facile si on pense à se ramener à la dimension 1. Le jury imagine bien que les candidats sont peu familiers avec les fonctions de plusieurs variables mais ils devraient au minimum savoir se ramener en dimension 1, soit en fixant toutes les variables sauf une, soit en considérant un segment. La partie (d) était plus difficile et n'a pas été vue : il fallait encore se ramener en dimension 1 et utiliser les inégalités de convexité usuelles.

Partie II Entropie et codage

Il y avait une confusion de notations : le log dans le texte aurait du être un ln. La plupart des candidats ont su détecter cette coquille afin que 4-b) soit vraie. Le jury a, bien évidemment, décidé de ne pas tenir rigueur aux candidats qui ont utilisé le logarithme en base 10 pourvu qu'il ait été convaincu que le candidat était capable de mener l'étude de fonctions en 4-b). Par ailleurs le poids de cette question a été réduit au minimum.

(4) Modulo la remarque précédente, la question a été très bien traitée.

(5) (a) a été bien traitée ; dans (b) il fallait veiller au mot vide ce qui a été peu fait ; avec l'indication (c) a été bien traité par une moitié des candidats : il s'agissait de raisonner par récurrence en veillant à l'initialisation.

(6) La question n'a pas posé de problème particulier sauf qu'une partie non négligeable des copies oublie d'invoquer 5-c) pour conclure.

Partie III Transport régularisé

(7) Il s'agissait de voir que $F(\alpha, \beta)$ était obtenu comme l'intersection du convexe Q avec un nombre fini d'hyper-espaces affines. Les candidats pour la plupart sont revenus à la définition de la convexité et ont ainsi perdu un peu de temps.

(8) Question de cours élémentaire, bien traitée par les candidats.

(9) Au final il fallait dire qu'une fonction linéaire est convexe plutôt que se lancer dans des calculs aveugles.

(10) (a) Autant l'aspect fermé a été fait efficacement, autant le côté borné a souvent été plus laborieux ; (b) bien traité dans les bonnes copies ; (c) peu vu sauf dans les très bonnes copies.

(11) (a) était plus difficile et n'a pas été vu ; (b) fait dans de rares copies : il ne faut pas hésiter à donner même un exemple trivial.

Partie IV Dualité

(12) (a) De nombreux candidats sont venus prendre un point facilement gagné ; (b) très peu de candidats sont parvenus à suivre l'indication et répondre à cette question finalement facile quand on parvient à garder en mémoire ce qui a été déjà fait dans le sujet.

(13) Le reste du sujet n'a pas été abordé, tout du moins correctement. En (a) il s'agissait de calculer les dérivées partielles $\partial_{q_{i,j}}\mathcal{L}$ et d'invoquer la stricte convexité de $\mathcal{L}$; (b) le calcul de $G(f,g)$ ne posait aucun problème, sauf à se perdre dans les notations, ce qui permettait alors dans (c) d'obtenir la concavité de G à partir de la convexité de la fonction exponentielle.

(14) Il suffisait de calculer $\partial_{f_i}G(f,g)$ et d'invoquer la concavité.

(15) Il fallait en fixant f à f^k (puis en fixant g à g^{k+1}) invoquer la concavité de $G(f^k,g)$ (resp. $G(f,g^{k+1})$) et utiliser que sa différentielle en g (resp. en f) s'annulait en $g^{k+1}=g_*(f^k)$ (resp. en $f^{k+1}=f_*(g^{k+1})$).

(16) (a) résulte de la continuité de f_* et g_* et de la concavité de G ; pour (b) on utilise que $(q(\epsilon),(f(\epsilon),g(\epsilon))$ est un point selle de $\mathcal{L}$.

Pour (c), on se ramène en dimension 1 en calculant la dérivée seconde de $G(f^{(t)},g^{(t)})$ où $f^{(t)} := (1-t)f(\epsilon)+tf^\infty$ et $g^{(t)} := (1-t)g(\epsilon)+tg^\infty$ et dire que puisque $G(f^{(0)},g^{(0)}) = G(f^{(1)},g^{(1)}) = G_*$ alors $G(f^{(t)},g^{(t)})$, qui est concave, n'est pas strictement convave, ce qui fournit la constante a demandée.

Enfin (d) découle alors directement des égalités $(f(\epsilon),g(\epsilon)) = (f^\infty + a, g^\infty - a)$ et $q(f,g) = q(f+a,g-a)$ et de la continuité de $(f,g)\mapsto q(f,g)$.

Le sujet ne posait pas de difficultés notables mais il fallait savoir, pour une fonction à plusieurs variables, se ramener en dimension 1 en considérant pour deux points x, y le segment $\{tx + (1-t)y : t \in [0,1]\}$. Seules les très bonnes copies ont su utiliser cette technique qui devrait être connue de tous les candidats. La difficulté essentielle de la dernière partie découlait de la lourdeur des objets considérés ce qui a visiblement effrayé les candidats : il est compréhensible qu'en fin d'épreuve il soit cependant plus difficile de se concentrer sur des énoncés chargés même si, habituellement, des notations fournies cachent rarement des arguments complexes.

Sur les premières parties du sujet qui ont été très largement abordées, les différences se sont faites sur la précision des arguments et des raisonnements : par exemple en 5-a) il fallait faire attention au mot vide, en 6) il fallait invoquer 5-c)...

Le jury a, comme à son habitude, utilisé tout l'échelle des notes, les meilleures copies étant notées à 20 et les plus mauvaises proches de 0. La moyenne a été fixée à 10 avec un écart type modeste de 3,5 reflétant la faible hétérogénéité des questions abordées. Les deux autres quartilles sont 7.6 et 12.6, dix pour cent des candidats ayant une note au dessus de 14,5.