Dualité lagrangienne pour la résolution d'un problème de décision avec contraintes

Edith Grall-Maës, Pierre Beauseroy

Institut Charles Delaunay (CNRS FRE 2848)
Laboratoire de Modélisation et Sûreté des Systèmes
Université de Technologie de Troyes
12, rue Marie Curie - BP 2060 -10010 Troyes cedex - FRANCE
Edith.Grall@utt.fr,Pierre.Beauseroy@utt.fr

Résumé – Le travail traite d'un problème de décision avec contraintes. Le problème consiste à déterminer la partition optimale d'un espace de représentation en régions, chacune d'elles étant associée à une décision, afin de minimiser un critère et de satisfaire des contraintes de performance. Il est démontré que la solution du problème dual est identique à celle du problème primal, garantissant ainsi l'absence de trou de dualité. Pour éviter de prouver la convexité de la fonction objective à minimiser et des contraintes, qui sont des fonctions de la partition, la démonstration est basée sur la propriété de convexité de la fonction de perturbation. Un exemple simulé illustre le problème traité et sa solution.

Abstract – This paper deals with a decision problem with constraints. The problem consists in determining the optimal partition of a representation space into regions, each of them defining a decision, in order to minimize a criterion and to verify performance constraints. It is shown that the solution of the dual problem is equal to the solution of the primal problem, guarantying the absence of duality gap. In order to bypass the difficulty of showing the convexity of both the objective function and the constraints, since they depend on a partition, this demonstration is based on the convexity of the perturbation function. A simulated example illustrates the problem and its solution.

1 Introduction

Les problèmes de décision sont généralement définis par leurs options de décision, qui correspondent aux classes du problème, et par un critère à optimiser. Toutefois dans de nombreuses applications, il est plus naturel de définir des exigences que doit respecter la règle de décision sous la forme de contraintes de performances à vérifier. Par exemple certaines applications de diagnostic médical, d'identification d'individus ou de détection de fraude nécessitent d'être conçues pour satisfaire des exigences précises de fausse alarme ou de probabilité d'erreur.

Dans le cadre des tests d'hypothèses statistiques, différentes règles de classification correspondent à la solution de problèmes avec contraintes de performance. Par exemple le test de Neyman-Pearson ([3]) correspond à la solution d'un problème avec contrainte sur la probabilité de fausse alarme, la règle de Chow ([2]) correspond à la solution d'un problème avec contrainte sur la probabilité d'erreur et la règle de Ha ([6]) correspond à la solution d'un problème avec contrainte sur le nombre moyen de classes sélectionnées dans un cadre multiclasses avec option de rejet sélectif de classes. Ces différents problèmes peuvent s'inscrire dans un formalisme général proposé dans [4, 5], permettant de décrire des problèmes de décision avec contraintes de performance variés. Dans ce formalisme, le problème est défini par les options de décision correspondant aux classes et sous-ensembles de classes d'affectation possibles, par les contraintes s'exprimant par des combinaisons linéaires de probabilités de décision conditionnelles aux classes, et par une fonction coût à minimiser. La solution recherchée est la règle de décision qui minimise la fonction coût tout en respectant les contraintes.

Ce formalisme peut être étendu en considérant des options de décision plus générales. Le problème consiste alors à déterminer la partition de l'espace de représentation en régions où chacune correspond à une décision, de façon à minimiser la fonction coût et respectant les contraintes données. Ce travail porte sur la résolution de ce type de problème, qui est un problème d'optimisation non linéaire avec contraintes. Des développements sont notamment menés pour démontrer que la partition optimale définissant la solution du problème dual est identique à la partition optimale définissant la solution du problème primal. Cela permet de montrer l'absence de trou de dualité.

Le paragraphe 2 est consacré à la définition du problème primal. Dans le paragraphe 3, le problème dual est exposé et dans le paragraphe 4, l'égalité entre la solution du problème dual et celle du problème primal est démontrée. Un exemple simulé illustre le problème et sa solution au paragraphe 5. Les conclusions sont données dans le dernier paragraphe.

2 Définition du problème

Le problème considéré est un problème de décision à N classes dans un espace de représentation $\mathbf{R}^d$, pour lequel toute règle de décision est définie par une partition Z de l'espace $\mathbf{R}^d$. Le problème est défini par trois caractéristiques : les options de décision, les contraintes, et la fonction coût.

Les options de décision définissent les décisions associées à chacune des régions de la partition. Une prise de décision consiste à affecter un élément x à une et une seule de ces options. L'option de décision sélectionnée pour l'élément x étant donnée la partition Z est notée $\delta^{(Z)}(x)$, et prend une valeur entière entre 1 et I correspondant au nombre d'options. Dans [4, 5], les options de décision correspondent aux différentes classes et à des sous-ensembles de classes. Affecter une observation à un sous-ensemble de classes revient à affirmer qu'elle appartient à l'une de ces classes sans spécifier laquelle. Ces options permettent donc d'introduire le rejet total et le rejet sélectif de classes.

Les contraintes, au nombre de K, sont définies par des combinaisons linéaires de probabilités de décision conditionnelles aux classes :

$$e^{(k)}(Z) \leq \gamma^{(k)}$$

où $\gamma^{(k)}$ est le seuil associé à l'expression de la contrainte k, et $e^{(k)}(Z)$ est l'expression de la contrainte k :

$$e^{(k)}(Z) = \sum_{j=1}^{N} \int_{R^n} \alpha^{(k)}_{\delta^{(Z)}(x),j} P(x/\omega_j) P_j dx. \qquad (1)$$

où $P(x/\omega_j)$ est la probabilité de x conditionnelle à la classe j et $\alpha^{(k)}_{\delta^{(Z)}(x),j}$ est le coût, dans la contrainte k, associé à la décision $\delta^{(Z)}(x)$ lorsque x appartient à la classe ω_j.

La fonction coût à minimiser $\overline{c}(Z)$ est définie par :

$$\overline{c}(Z) = \sum_{j=1}^{N} \int_{R^n} c_{\delta^{(Z)}(x),j} P(x/\omega_j) P_j dx \qquad (2)$$

où $c_{\delta^{(Z)}(x),j}$ est le coût associé à la décision $\delta^{(Z)}(x)$ lorsque x appartient à la classe ω_j dans la fonction coût.

Étant données ces caractéristiques, le problème consiste à déterminer la partition qui minimise la fonction coût et vérifie les K contraintes :

$$\begin{aligned} &\min_Z \ \overline{c}(Z) \\ &\text{sous les contraintes } \ e^{(k)}(Z) \leq \gamma^{(k)} \ \forall k = 1..K. \end{aligned} \qquad (3)$$

3 Problème dual

Le problème dual associé au problème primal défini en section 2 est défini par ([7]) :

$$\begin{aligned} &\max_\mu \{w(\mu)\} \\ &\text{sous les contraintes } \ \mu \in \mathbf{R}^{K+} \end{aligned} \qquad (4)$$

où w est la fonction duale :

$$w(\mu) = \inf_Z L(Z, \mu)$$

et $L(Z, \mu)$ est la fonction de Lagrange :

$$\begin{aligned} L(Z,\mu) &= \overline{c}(Z) + \sum_{k=1}^{K} \mu_k \left(e^{(k)}(Z) - \gamma^{(k)} \right) \\ &= \overline{c}(Z) + \mu^T e(Z) - \mu^T \gamma \end{aligned} \qquad (5)$$

où $\mu = [\mu_1, \mu_2 \ldots \mu_K]$ est le vecteur des coefficients de Lagrange associés à chacune des contraintes, et $\gamma = [\gamma_1, \gamma_2 \ldots \gamma_K]$ est le vecteur des seuils de contraintes.

Compte tenu de (1) et (2), pour une valeur de μ donnée, la partition qui minimise $L(Z, \mu)$ est donnée par

$$\widetilde{Z_i}(\mu) = \{x | \lambda_i(x, \mu) < \lambda_l(x, \mu), l = 1..I, \ l \neq i\}$$

avec $\lambda_i(x, \mu)$ défini par :

$$\lambda_i(x, \mu) = \sum_{j=1}^{N} P_j P(x/\omega_j) \left(c_{i,j} + \sum_{k=1}^{K} \mu_k \alpha^{(k)}_{i,j} \right).$$

Donc la résolution du problème dual se ramène à la résolution d'un problème simple, consistant à maximiser :

$$w(\mu) = L(\widetilde{Z}(\mu), \mu).$$

Le maximum obtenu pour μ^* est associé à la partition optimale Z^*.

Néanmoins la solution du problème dual est égale à celle du problème primal seulement s'il n'y pas de trou de dualité. Il est donc nécessaire de vérifier l'absence de trou de dualité pour garantir que le problème primal peut être résolu à l'aide du lagrangien.

4 Égalité des solutions du problème primal et du problème dual

Pour montrer que la valeur optimale du problème dual est égale à celle du problème primal, généralement, le théorème de la dualité forte est utilisé. Néanmoins cela implique l'étude de la convexité de la fonction coût est des fonctions de contrainte. Or pour ce problème, ces fonctions dépendent d'une partition, ce qui rend la convexité difficile à montrer. C'est pourquoi l'égalité est plutôt montrer en faisant appel aux deux théorèmes suivants ([1, 7]). Selon le premier, une condition suffisante pour que Z^* soit la valeur optimale du problème primal (3) est que (Z^*, μ^*) soit un point selle de la fonction lagrangienne (5). Selon le second, dans l'hypothèse où le problème primal admet une solution, (Z^*, μ^*) est un point selle de la fonction lagrangienne si et seulement si

$$v(y) \geq v(\gamma) - \mu^{*T}(y - \gamma) \quad \forall y = [y^{(1)}, y^{(2)} \ldots y^{(K)}] \in \mathbf{R}^K \qquad (6)$$

où $v : \mathbf{R}^K \to \mathbf{R}$ est la fonction de perturbation définie par :

$$v(y) = \min_Z \left\{ \overline{c}(Z) | e^{(k)}(Z) \leq y^{(k)} \ \forall k = 1 \ldots K \right\}.$$

Si la fonction de perturbation est une fonction convexe propre et si $y = \gamma$ appartient à l'intérieur du domaine de la fonction, alors la fonction admet un sous-gradient au point $y = \gamma$ et la condition (6) est vérifiée. Or quand une solution au problème primal existe le point $y = \gamma$ appartient à l'intérieur du domaine de la fonction sauf pour la cas particulier où il serait exactement sur le bord du domaine, ce qui est, d'un point de vue probabiliste, impossible. Par conséquent si la fonction de perturbation est propre convexe, alors les problèmes primal et dual ont la même valeur optimale, et cela garantit l'absence de trou de dualité.

Montrons que v est propre convexe selon la définition :

$$v\left(\beta y_1 + (1-\beta)y_2\right) \leq \beta v(y_1) + (1-\beta)v(y_2)$$

où $\beta \in [0,1]$.

Étant donnés deux vecteurs y_1 et y_2, notons $Z^{(1)}$ et $Z^{(2)}$ les partitions correspondant respectivement à $v(y_1)$ et $v(y_2)$, alors

$$v(y_1) = \overline{c}(Z^{(1)}) \text{ et } v(y_2) = \overline{c}(Z^{(2)}) \qquad (7)$$
$$e^{(k)}(Z^{(1)}) \leq y_1^{(k)}\ \forall k \text{ et } e^{(k)}(Z^{(2)}) \leq y_2^{(k)}\ \forall k.$$

Puisque $v(y)$ est une fonction décroissante, nous obtenons :

$$v\left(\beta y_1 + (1-\beta)y_2\right) \leq v\left(\beta e(Z^{(1)}) + (1-\beta)e(Z^{(2)})\right).$$

Considérons R^n comme un ensemble d'hypercubes de côtés h : $w_{m_1,m_2,\ldots,m_n} = [m_1 h, (m_1+1)h] \times [m_2 h, (m_2+1)h]] \times \ldots [m_n h, (m_n+1)h]$. Définissons $E^{(k)}(Z^{(i)})$ ($i=1$, or 2) par :

$$E^{(k)}(Z^{(i)}) = \sum_{j=1}^{N} \sum_{m_1,\ldots,m_n} \alpha^{(k)}_{\delta_{Z^{(i)}}(x_{m_1,\ldots,m_n}),j} P(x_{m_1,\ldots,m_n}/\omega_j)P_j h^n$$

où $x_{m_1,\ldots,m_n}$ est un point de $w_{m_1,\ldots,m_n}$, et définissons $\overline{C}(Z^{(i)})$ de façon similaire :

$$\overline{C}(Z^{(i)}) = \sum_{j=1}^{N} \sum_{m_1,\ldots,m_n} c_{\delta_{Z^{(i)}}(x_{m_1,\ldots,m_n}),j} P(x_{m_1,\ldots,m_n}/\omega_j)P_j h^n.$$

Dans l'hypothèse où les densités de probabilité sont continues, d'après (1) et (2) nous obtenons :

$$\lim_{h\to 0} E^{(k)}(Z^{(i)}) = e^{(k)}(Z^{(i)}) \forall k$$

et

$$\lim_{h\to 0} \overline{C}(Z^{(i)}) = \overline{c}(Z^{(i)}).$$

Définissons une partition $Z^{(\beta)}$ en divisant chaque hypercube en deux hypercubes $w^{(1)}_{m_1,\ldots,m_n}$ et $w^{(2)}_{m_1,\ldots,m_n}$ de volumes respectifs βh^n et $(1-\beta)h^n$, et en définissant les décisions d'affection selon :

$$\forall x \in w^{(1)}_{m_1,\ldots,m_n}, \delta_{Z^{(\beta)}}(x) = \delta_{Z^{(1)}}(x_{m_1,\ldots,m_n})$$
$$\text{et } \forall x \in w^{(2)}_{m_1,\ldots,m_n}, \delta_{Z^{(\beta)}}(x) = \delta_{Z^{(2)}}(x_{m_1,\ldots,m_n}).$$

En définissant $E^{(k)}(Z^{(\beta)})$ selon :

$$E^{(k)}(Z^{(\beta)}) =$$
$$\sum_{j=1}^{N} \sum_{m_1,\ldots,m_n} \alpha^{(k)}_{\delta_{Z^{(1)}}(x_{m_1,\ldots,m_n}),j} P(x_{m_1,\ldots,m_n}/\omega_j)P_j \beta h^n$$
$$+\sum_{j=1}^{N} \sum_{m_1,\ldots,m_n} \alpha^{(k)}_{\delta_{Z^{(2)}}(x_{m_1,\ldots,m_n}),j} P(x_{m_1,\ldots,m_n}/\omega_j)P_j (1-\beta) h^n$$

et en prenant la limite quand h tend vers 0 nous obtenons

$$e^{(k)}(Z^{(\beta)}) = \beta e^{(k)}(Z^{(1)}) + (1-\beta)e^{(k)}(Z^{(2)}) \quad \forall k \qquad (8)$$

et

$$\overline{c}(Z^{(\beta)}) = \beta \overline{c}(Z^{(1)}) + (1-\beta)\overline{c}(Z^{(2)}). \qquad (9)$$

Soit la partition $Z^{(\beta)*}$ telle que (8) soit vérifiée et telle que $\overline{c}$ soit minimium, alors

$$v\left(\beta e(Z^{(1)}) + (1-\beta)e(Z^{(2)})\right) = \overline{c}(Z^{(\beta)*}) \leq \overline{c}(Z^{(\beta)}). \qquad (10)$$

Á partir de (4), (10), (9) et (7) il s'ensuit que

$$v\left(\beta y_1 + (1-\beta)y_2\right) \leq \beta v(y_1) + (1-\beta)v(y_2)$$

ce qui prouve que v est convexe.

En outre, puisque $v(y)$ est supérieur ou égal au coût sans contraintes, le domaine de v est non vide et $v(y) > -\infty$ pour tout y dans le domaine de v. Cela complète la preuve et permet de conclure que v est propre convexe.

5 Exemple

Soit un problème de classification dans $\mathbf{R}^2$ avec 3 classes équiprobables de distributions normales, dont les moyennes et les matrices de variance-covariance sont définies par : $m_1 = (-1.1; 0), \Sigma_1 = I, m_2 = (1.1; 0)$, $\Sigma_2 = I$, $m_3 = (0; 2), \Sigma_1 = 0.5I$ où I est la matrice identité. Les densités de probabilités sont représentées sur la figure 1 par des courbes de niveau.

Les caractéristiques du problème sont définies par :

- 7 options de décision : $\{1\}, \{2\}, \{3\}, \{1;2\}, \{1;3\}, \{2;3\}, \{1;2;3\}$, par exemple, quand un élément est affecté à l'option $\{1;2\}$ cela signifie qu'il est affecté à la classe 1 ou à la classe 2 et qu'il y a ambiguïté
- 2 contraintes :
$$\begin{cases} P_E \leq 0.05 \\ P_I \leq 0.08 \end{cases}$$
  où P_E est la probabilité d'erreur :
$$P_E = \frac{1}{3}\left(P(D_2/\omega_1) + P(D_3/\omega_1) + P(D_6/\omega_1) + P(D_1/\omega_2) + P(D_3/\omega_2) + P(D_5/\omega_2) + P(D_1/\omega_3) + P(D_2/\omega_3) + P(D_4/\omega_3)\right)$$
  et P_I la probabilité d'indécision (classification correcte mais avec de l'ambiguïté) :
$$P_I = \frac{1}{3}\left(P(D_4/\omega_1) + P(D_4/\omega_2) + P(D_5/\omega_1) + P(D_5/\omega_3) + P(D_6/\omega_2) + P(D_6/\omega_3)\right)$$
- le coût moyen donné par : $c = P_E + 0,5P_I + P_R$ avec P_R la probabilité de non décision (usuellement appelé rejet) : .
$$P_R = \frac{1}{3}\left(P(D_7/\omega_1) + P(D_7/\omega_2) + P(D_7/\omega_3)\right).$$
  La fonction coût correspond à un coût de 1 pour une classification erronée ou pour aucune décision, et un coût de 0,5 pour une classification correcte avec ambiguïté.

La partition optimale obtenue après résolution du problème dual est représentée sur la figure 1.

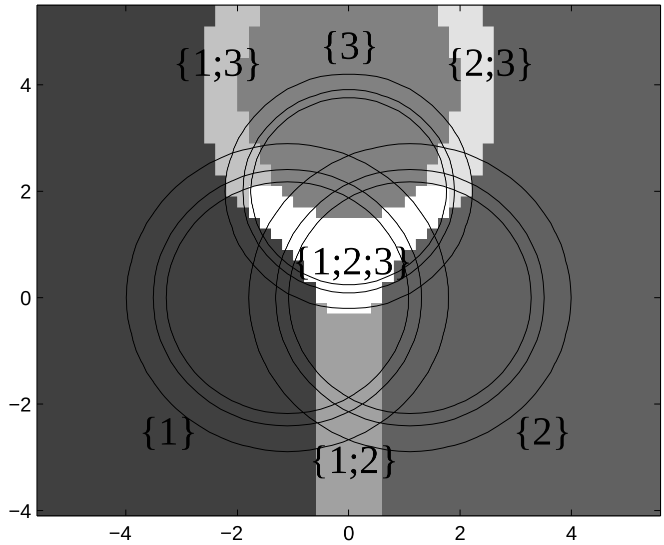

FIG. 1 – Densités de probabilités et partition optimale

6 Conclusion

L'étude du problème dual associé à un problème d'optimisation de la partition de l'espace de représentation en présence de contraintes a permis de démontrer que la solution du problème dual est toujours égale à la solution du problème primal : il n'y a jamais de trou de dualité. L'égalité entre les deux solutions s'appuie généralement sur la propriété de convexité de la fonction à minimiser et des fonctions définissant les contraintes. Dans le problème considéré, ces fonctions ayant pour paramètre une partition, il est difficile de montrer la convexité de ces fonctions. L'égalité entre les deux solutions a été prouvée par l'étude de la convexité de la fonction de perturbation.

Par conséquent, dans l'hypothèse où le problème primal admet une solution, cette solution pourra toujours être déterminée par la résolution du problème dual et, le problème initial avec contraintes admet un problème équivalent sans contraintes.

La formulation du problème étant très générale, la solution d'une grande variété de problèmes de décision avec contraintes peuvent être résolus aisément dans le cadre des tests d'hypothèses statistiques. Des travaux actuels consistent à traiter ce type de problèmes dans le cadre de l'apprentissage supervisé.

Références

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty *Nonlinear Programming : Theory and Algorithms.* Wiley, 1993.

[2] C. Chow, On optimum recognition error and reject tradeoff, IEEE Transactions on Information Theory IT-16 (1) (1970) 41–46.

[3] K. Fukunaga, Introduction to Statistical Pattern Recognition, Boston, 1990.

[4] E. Grall-Maës, P. Beauseroy and A. Bounsiar "Multilabel Classification Rule with Performance Constraints ," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May, 2006.

[5] E. Grall-Maës, P. Beauseroy, and A. Bounsiar, "Quality assessment of a supervised multilabel classification rule with performance constraints," in *14th European Signal Processing Conference (EUSIPCO'06)*, September, 2006.

[6] T. Ha, The optimum class-selective rejection rule, Transactions on Pattern Analysis and Machine Intelligence 19 (6) (1997) 608–615.

[7] M. Minoux, *Mathematical Programming : Theory and Algorithms.* Wiley, 1986.