Optimisation sous contraintes

Soit le problème d'optimisation

$$(P) : \begin{array}{ll} \min & f(x) \\ \text{s.c.} & h_i(x) \leq 0 \\ & g_i(x) \geq 0 \\ & m_i(x) = 0 \end{array}$$

Lagrangien ou fonction de Lagrange associé à ce problème

$$L(x, \alpha, \theta, \lambda) = f(x) - \sum_i \lambda_i h_i(x) - \sum_i \alpha_i g_i(x) - \sum_i \theta_i m_i(x)$$

λ_i ($\in \mathbb{R}_+$) positif, α_i négatif et $\theta_i \in \mathbb{R}$ pour tout i

- Deux conditions à vérifier pour résoudre ce problème, lesquelles conditions sont similaires à

→ La dérivée de L est nulle ($L'(x) = 0$ permet de trouver l'extremum)

→ $L''(x) > 0$ permet de prouver que cet extremum est un minimum.

Condition du 1er ordre

$\nabla L(x, \lambda, \alpha, \theta) = 0$, ($\nabla L(x_i) = \left(\frac{\partial L}{\partial x_i}\right)_i$)

permet de déterminer $x^*, \lambda^*, \alpha^*, \theta^*$

Condition du 2nd ordre

$HL \geq 0$,

La matrice hessienne de L est semi-définie positive, i.e.

($v^T (HL) v \geq 0$ pour tout vecteur v non nul)

Exemple 1 : Calcul du gradient d'une fonction

$$g((x,y),\lambda) = x^3 + y - \lambda(1 - x^2 - 2y^2)$$

Solution :

$$\nabla g((x,y),\lambda) = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial \lambda}\right)$$

est le vecteur constitué des dérivées premières de g par rapport à chacune de ses composantes (x, y et λ).

$$\frac{\partial g}{\partial x} = 3x^2 + 2\lambda x \quad , \quad \frac{\partial g}{\partial y} = 1 + 4\lambda y.$$

$$\frac{\partial g}{\partial \lambda} = x^2 + 2y^2 - 1.$$

Exemple 2 : Calcul de la matrice Hessienne d'une fonction

Calculons la matrice hessienne de g.

$$H_g = \begin{bmatrix} \partial^2 g/\partial x^2 & \partial^2 g/\partial x \partial y & \partial^2 g/\partial x \partial \lambda \\ \partial^2 g/\partial y \partial x & \partial^2 g/\partial y^2 & \partial^2 g/\partial y \partial \lambda \\ \partial^2 g/\partial \lambda \partial x & \partial^2 g/\partial \lambda \partial y & \partial^2 g/\partial \lambda^2 \end{bmatrix}$$

$$\frac{\partial^2 g}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial g}{\partial x}\right) = 6x + 2\lambda.$$

De même,

$$\frac{\partial^2 g}{\partial y^2} = 4\lambda \quad ; \quad \frac{\partial^2 g}{\partial \lambda^2} = 0 \quad ;$$

$$\frac{\partial^2 g}{\partial x \partial y} = \frac{\partial^2 g}{\partial y \partial x} = 0 \quad ; \quad \frac{\partial^2 g}{\partial x \partial \lambda} = \frac{\partial^2 g}{\partial \lambda \partial x} = 2x$$

$$\frac{\partial^2 g}{\partial y \partial \lambda} = \frac{\partial^2 g}{\partial \lambda \partial y} = 4y.$$

$$H_g = \begin{bmatrix} 6x + 2\lambda & 0 & 2x \\ 0 & 4\lambda & 4y \\ 2x & 4y & 0 \end{bmatrix}$$

Remarque : La matrice Hessienne de toute fonction est une matrice symétrique

Exemple 3 : Lagrangien d'un problème d'optimisation avec contraintes.

Considérons le problème de minimisation

$$(P) : \min J(x,y) := x^3 + y$$
$$\text{s.c.} \quad x^2 + 2y^2 - 1 = 0 , \quad x \geq 0$$

Le lagrangien de cette fonction est :

$$L(x,y,\lambda) = x^3 + y - \lambda(x^2 + 2y^2 - 1) \quad , \quad \lambda \in \mathbb{R}.$$

Exemple 4 : Résoudre le problème d'optimisation de l'exemple 3.

Solution :

→ Condition de 1er ordre :

$$\nabla L(x,y,\lambda) = 0 \iff \begin{cases} 3x^2 + 2\lambda x = 0 \\ 1 + 4\lambda y = 0 \\ x^2 + 2y^2 - 1 = 0. \end{cases}$$

$$\iff \begin{cases} x(3x + 2\lambda) = 0 \\ 1 + 4\lambda y = 0 \\ x^2 + 2y^2 = 1 \end{cases}$$

d'où

$$x^* = \{-2\lambda/3 , 0\} \text{ (car } x \geq 0 \text{)}$$
$$y^* = -1/4\lambda$$
$$\lambda^* \in \mathbb{R}_- \ (\lambda \leq 0) \quad \text{car } x \geq 0.$$

→ Condition de 2nd ordre.

Il s'agit juste de vérifier que si

$$Hf(x^*, y^*, \lambda^*) \succcurlyeq 0$$

pour conclure que (x^*, y^*) est une solution à ce problème.

Exemple 5. Comment montrer qu'une matrice est semi-définie positive ou non ?

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Soit $v \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2 \setminus \{(0,0)\}$

$$v^T A v = (v_1 \;\; v_2) \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

$$= 2v_1^2 + v_2^2 \geq 0$$

Donc $A \succcurlyeq 0$ (semi définie positive)

Application : Estimation des paramètres d'un modèle de régression linéaire sous contraintes.

On veut estimer par la méthode des moindres carrés ordinaires le paramètre β du modèle de régression

$$Y = X\beta + \varepsilon.$$

sous la contrainte $R\beta = 0$

Ceci revient à minimiser la somme des carrés ~~résiduel~~ residuel sous cette même contrainte. D'où le problème d'optimisation

$$\min \sum_i \varepsilon_i^2 \quad \text{s.c.} \quad R\beta = c \iff \min \|\varepsilon\|^2 \quad \text{s.c.} \quad R\beta = c$$

$$\iff \min_\beta \|Y - X\beta\|^2 \quad \text{s.c.} \quad R\beta - c = 0$$

Le lagrangien de ce problème est donné par:

$$L(\beta, \lambda) = \|Y - X\beta\|^2 - \lambda(R\beta - c)$$

→ <u>Condition de 1er ordre</u>, $\nabla L(\beta, \lambda) = 0$

$$\nabla L(\beta, \lambda) = \left(\frac{\partial L}{\partial \beta}, \frac{\partial L}{\partial \lambda} \right).$$

$$\frac{\partial L}{\partial \beta} = -2X^T(Y - X\beta) - \lambda R$$

$$= -2X^TY + 2X^TX\beta - \lambda R.$$

$$\frac{\partial L}{\partial \lambda} = -(R\beta - c).$$

$$\nabla L(\beta, \lambda) = 0 \iff \begin{cases} -2X^TY + 2X^TX\hat{\beta}_c - \hat{\lambda} R = 0 & (1) \\ R\hat{\beta}_c - c = 0 & (2) \end{cases}$$

Multiplier (1) par $R(X^TX)^{-1}$ et obtenir

$$-2R(X^TX)^{-1}X^TY + 2R\underbrace{(X^TX)^{-1}X^TX}_{I_p}\hat{\beta}_c - R(X^TX)^{-1}R^T\hat{\lambda} = 0$$

$$-2R(X^TX)^{-1}X^TY + 2R\hat{\beta}_c - R(X^TX)^{-1}R^T\hat{\lambda} = 0$$

$\left(\text{Car } (X^TX)^{-1}(X^TX) = I_p\right)$

$$-2R(X^TX)^{-1}X^TY + 2c - R(X^TX)^{-1}R^T\hat{\lambda} = 0$$

$\left(\text{Car } R\hat{\beta}_c = c \text{ d'après } (2)\right)$

Nous obtenons alors pour $\hat{\lambda}$

$$\hat{\lambda} = 2\left[R(X^TX)^{-1}R^T\right]^{-1}\left[c - R(X^TX)^{-1}X^TY\right]$$

Remplaçons ensuite $\hat{\lambda}$ dans la 1ère équation :

$$-2X^TY + 2X^TX\hat{\beta}_c - 2R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left[c - R(X^TX)^{-1}X^TY\right] = 0$$

D'où

$$\hat{\beta}_c = (X^TX)^{-1}X^TY + (X^TX)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}(c - R\hat{\beta})$$

$$\boxed{\hat{\beta}_c = \hat{\beta}_{MCO} + (X^TX)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}(c - R\hat{\beta}).}$$

N.B : La fonction $\|Y - X\beta\|_2^2$ étant convexe, $\hat{\beta}_c$ est l'unique solution du problème d'optimisation. Nous n'avons donc plus besoin de vérifier la condition de 2nd ordre.

Exemple: Un expérimentateur aimerait représenter le modèle

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

par la méthode des moindres carrés ordinaires, sous la contrainte $\beta_1 = 1$.

Pourrait-il le faire par le modèle

$$Y - X_1 = \beta_0 + \beta_2 X_2 + \varepsilon \; ?$$

Solution: Oui

Il pourrait procéder comme suit :

- Poser $\tilde{Y} = Y - X_1$. Alors
- $\tilde{Y} = \beta_0 + \beta_2 X_2 + \varepsilon$.
- Estimer $\hat{\beta}$ par MCO et obtenir

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_2 \end{pmatrix}$$

- La solution est : $\hat{\beta}_c = \begin{pmatrix} \hat{\beta}_0 \\ 1 \\ \hat{\beta}_2 \end{pmatrix}$

Autre méthode: Utilisant le résultat précédent, on pourrait remarquer que pour

$$R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \text{ et } c = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$R\beta = c \iff \beta_1 = 1.$$

La solution $\hat{\beta}_c$ est donc

$$\hat{\beta}_c = \hat{\beta} + (X^T X)^{-1} R^T \left[R (X^T X)^{-1} R^T \right]^{-1} (c - R\hat{\beta})$$

où $\hat{\beta}$ est le paramètre estimé de β par la méthode des moindres carrés ordinaires sans aucune contrainte.

Partie 2 : « Dummy Variables »

Ces variables interviennent lorsqu'on veut représenter par un modèle des données provenant de plusieurs groupes.

Illustration : 2 groupes

i	Groupe 1 y_i	x_i	Groupe 2 y'_i	x'_i
1	y_1	x_{11}	y'_1	x'_{11}
2	y_2	x_{12}	y'_2	x'_{22}
3	y_3	x_{13}	y'_3	x'_{33}

Afin de représenter ces données par un modèle de régression, on fera intervenir la variable "dummy" qui code 1 si on est dans le groupe 1 et 0 sinon (ou inversement).
On a donc :

$$Y = X\beta + \varepsilon.$$

où

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y'_1 \\ y'_2 \\ y'_3 \end{bmatrix} \qquad X = \begin{array}{c} \begin{array}{ccc} 1 & x & z \end{array} \\ \begin{bmatrix} 1 & x_{11} & 1 \\ 1 & x_{12} & 1 \\ 1 & x_{13} & 1 \\ \hline 1 & x_{11} & 0 \\ 1 & x'_{12} & 0 \\ 1 & x'_{13} & 0 \end{bmatrix} \end{array} \begin{array}{l} \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} \text{Groupe 1} \\ \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} \text{Groupe 2} \end{array}$$

En pratique, la matrice de design X contient aussi une variable d'interaction entre x et z ("dummy variable")

N.B : p groupes donnent lieu à p-1 « dummy variable »

Exemple :

Groupe A		Groupe B	
X	Y	X	Y
8	5.3	9	5.1
0	0.9	7	4.4
12	7.1	8	5.2
2	2.4	6	3.8

Soit la variable « dummy »

$$Z = \begin{cases} 0 & \text{si groupe A} \\ 1 & \text{sinon.} \end{cases}$$

Alors

$$Y = \beta_0 + \beta_1 X + \alpha_0 Z + \alpha_1 \underbrace{XZ}_{\text{Interaction entre X et Z}} + \varepsilon$$

$$Y^T = (5.3, 0.9, 7.1, 2.4, 5.1, 4.4, 5.2, 3.8)$$

$$X = (8, 0, 12, 2, 9, 7, 8, 6)$$

$$Z = (0, 0, 0, 0, 1, 1, 1, 1)$$

$$T := XZ = (0, 0, 0, 0, 9, 7, 8, 6)$$

$$Y = \beta_0 + \beta_1 X + \alpha_0 Z + \alpha_1 T + \varepsilon$$

L'implémenter dans R et obtenir

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\alpha}_0, \hat{\alpha}_1) = (1.142, 0.506; -0.0418; -0.0360)$$

On pourrait effectuer des tests de significativité des variables Z et $T := XZ$ i.e

$H_0 : \alpha_0 = \alpha_1 = 0$ & $H_a : \alpha_0 \neq 0$ ou $\alpha_1 \neq 0$

Le modèle sans Z et T est :

$$\hat{Y} = 1.075 + 0.492 X$$

Statistique de Fisher (pour comparer ces 2 modèles)

1er modèle : $\hat{Y} = 1.142 + 0.586 X - 0.0418 Z - 0.036 XZ$

2nd modèle : $\hat{Y} = 1.075 + 0.492 X$

$$F = \frac{0.1818/2}{0.3272/4} = 1.11 \sim \mathcal{F}_{2,4}(\alpha)$$

seuil.

Au seuil $\alpha = 5\%$ par exemple,

$$F > \mathcal{F}_{2,4}(\alpha)$$

Considérer donc le modèle 2.