

Conditions nécessaires de Kuhn et Tucker en « algèbre géométrique ».

Dans la note « Expression covariante de l'optimisation sous contrainte ... » nous avons mis l'accent sur une démonstration tensorielle très simple des conditions de Kuhn et Tucker, qui évite en particulier de faire appel au fastidieux lemme de Farkas, et qui respecte le caractère en général non métrique des espaces considérés.

Nous allons faire ici exactement le contraire en utilisant les possibilités offertes par l'algèbre géométrique dans un espace muni d'une forme quadratique.

Considérons donc le programme optimal défini par :

$$I \quad \begin{cases} \text{Max } f(x) \\ g^k(x) \geq 0 \\ k \in (1, 2, \dots, p) \end{cases} \quad x \in E^n$$

Soient x_0 un optimum local, $g^s(x)$ les contraintes saturées en x_0 , en nombre m .

Nous supposons que nous étudions le voisinage d'un point « normal », c'est à dire où les gradients $\nabla g^s(x_0)$ sont linéairement indépendants et donc aussi non nuls. Bien entendu $g^s(x_0) = 0$.

Soient ε un scalaire positif, a une direction vectorielle. Une petite variation $\delta g^s(x_0)$ des contraintes saturées s'exprime par :

$$(1) \quad \delta g^s(x_0) = g^s(x_0 + \varepsilon a) - g^s(x_0) = \varepsilon a \cdot \nabla g^s(x_0) + \varepsilon o(\varepsilon)$$

Donc :

$$(2) \quad a \cdot \nabla g^s(x_0) \geq 0 \quad \text{si } a \text{ est une direction admissible, c'est à dire telle que l'on}$$

puisse trouver ε suffisamment petit pour que $(x_0 + \varepsilon a)$ satisfasse aux contraintes.

De même on a :

$$(3) \quad \delta f(x_0) = f(x_0 + \varepsilon a) - f(x_0) = \varepsilon a \cdot \nabla f(x_0) + \varepsilon o(\varepsilon)$$

Au voisinage d'un point normal des directions admissibles existent (démonstration par l'absurde).

Pour que x_0 soit un optimum local **il faut** que l'on ait :

$$\delta f(x_0) \leq 0$$

et donc :

$$(4) \quad a \cdot \nabla f(x_0) \leq 0$$

Les $\nabla g^s(x_0)$ définissent un sous-espace \mathcal{A}_m de E^n , dont le complément orthogonal est \mathcal{CA}_m .

On peut alors décomposer les vecteurs selon ces deux sous-espaces :

$$\begin{aligned} \nabla f(x_0) &= \nabla f(x_0)_{\parallel} + \nabla f(x_0)_{\perp} = -\lambda_s \nabla g^s(x_0) + \nabla f(x_0)_{\perp} \\ a &= a_{\parallel} + a_{\perp} \end{aligned}$$

et l'on obtient :

$$(5) \quad a \cdot \nabla f(x_0) = -\lambda_s a \cdot \nabla g^s(x_0) + a_{\perp} \cdot \nabla f(x_0)_{\perp}$$

Compte tenu du fait que a_{\perp} peut être choisi librement dans \mathcal{CA}_m la réalisation de (4) impose :

$$\nabla f(x_0)_{\perp} = 0 \quad \nabla f(x_0) \in \mathcal{A}_m$$

Il reste donc :

$$(6) \quad a \cdot \nabla f(x_0) = -\lambda_s a \cdot \nabla g^s(x_0)$$

Considérons pour un i_0 particulier, le vecteur réciproque a_{i_0} défini par les relations :

$$(7) \quad a_{i_0} \cdot \nabla g^s(x_0) = \delta_{i_0}^s \quad a_{i_0||} = a_{i_0} \in \mathcal{A}_m$$

Donc :

$$(8) \quad -\lambda_s \delta_{i_0}^s = -\lambda_{i_0} = a_{i_0} \cdot \nabla f(x_0) = \lim_{a \rightarrow a_{i_0}} a \cdot \nabla f(x_0) \leq 0$$

Donc il faut que tous les λ_s soient positifs.

Le passage à la limite de la relation (8) est celui que l'on trouve dans les exposés habituels en définissant au voisinage de x_0 une trajectoire admissible qui peut être tangente au cône tangent, notion que nous n'avons pas utilisée ici.

On retrouve bien sûr l'interprétation économique des coefficients λ_s puisque l'on peut écrire, d'après (3) et (8) :

$$(9) \quad -\varepsilon \lambda_s = \varepsilon a_s \cdot \nabla f(x_0) = f(x_0 + \varepsilon a_s) - f(x_0) + \varepsilon o(\varepsilon)$$

Nous avons en définitive établi les conditions nécessaires de Kuhn et Tucker, étant entendu que l'on associera des multiplicateurs nuls aux contraintes non saturées en x_0 :

$$\text{II} \quad \left\{ \begin{array}{l} \nabla f(x_0) + \lambda_k \nabla g^k(x_0) = 0 \\ \lambda_k g^k(x_0) = 0 \\ g^k(x_0) \geq 0 \quad \lambda_k \geq 0 \end{array} \right.$$

On voit que cette démonstration est encore bien plus courte que la démonstration tensorielle, tout simplement parce que l'on fait l'économie de l'introduction non seulement d'un, mais de deux systèmes de coordonnées, et de la loi de transformation.

Il est intéressant de noter en passant que, si a_s est l'un quelconque des vecteurs réciproques définis par (7), on obtient :

$$g^s(x_0 + \varepsilon a_s) = \varepsilon + \varepsilon o(\varepsilon)$$

Les vecteurs réciproques s'identifient donc aux vecteurs de base (en x_0) du système de coordonnées curvilignes utilisé dans la méthode tensorielle. Il revient au même de noter qu'un vecteur a_s donné est, d'après les relations (7), tangent à toutes les surfaces représentant les contraintes autres que g^s .

Le seul inconvénient, mais ici il est minime, est que l'on est obligé de superposer à un espace a priori affine un espace métrique R^n par un isomorphisme. Mais celui-ci ne joue qu'un rôle d'intermédiaire de calcul, qui disparaît à la fin puisque les équations (II) ont un caractère affine. En revanche dans l'étude tensorielle les calculs ne peuvent être faits que directement en espace affine.

Toute cette démonstration ne nécessite pas à vrai dire d'exploiter à fond les possibilités offertes par l'algèbre géométrique. Les propriétés standard de l'algèbre linéaire seraient suffisantes, à la réserve près que la GA permet une définition immédiate et simple des vecteurs réciproques¹, dits duals dans le vocabulaire usuel. D'autre part la GA met à notre avis plus clairement en évidence les propriétés géométriques de l'espace que les autres méthodes, sans doute parce qu'elle est plus concrète.

Ajoutons enfin que l'on peut facilement étendre les conditions K-T aux cas où plus de n contraintes sont saturées en x_0 . Il suffit d'en particulariser n et d'ajouter les contraintes supplémentaires une à une en vérifiant qu'elles laissent subsister un domaine admissible. Les vecteurs réciproques sont très utiles dans ce genre de démonstration, car tangents au domaine admissible définis par les n premiers vecteurs.

1. Le lecteur intéressé pourra vérifier cela dans le cours de Cambridge sur Internet.

Complément sur les conditions suffisantes du second ordre.

L'intérêt de l'utilisation d'une méthode hors coordonnées apparaît encore plus nettement dans une étude rapide des conditions suffisantes pour un optimum local.

Soit le lagrangien

$$\text{III} \quad \mathcal{L}(x) = f(x) + \lambda_k g^k(x) = f(x) + \lambda_s g^s(x)$$

À l'optimum nous avons :

$$(10) \quad \mathcal{L}(x_0) = f(x_0) \quad a.\nabla \mathcal{L}(x_0) = 0$$

Donc :

$$(11) \quad \mathcal{L}(x_0 + \varepsilon a) - \mathcal{L}(x_0) = f(x_0 + \varepsilon a) - f(x_0) + \lambda_s g^s(x_0 + \varepsilon a) = \frac{1}{2} \varepsilon^2 (a.\nabla)^2 \mathcal{L}(x_0) + \varepsilon^2 o(\varepsilon)$$

$$(12) \quad f(x_0 + \varepsilon a) - f(x_0) = -\lambda_s g^s(x_0 + \varepsilon a) + \frac{1}{2} \varepsilon^2 (a.\nabla)^2 \mathcal{L}(x_0) + \varepsilon^2 o(\varepsilon)$$

Bien entendu nous choisissons $(x_0 + \varepsilon a)$ de telle manière que les contraintes soient respectées. D'autre part nous distinguons parmi les indices s ceux, σ , pour lesquels les λ_σ sont strictement positifs, et ceux, $\bar{\sigma}$, pour lesquels les $\lambda_{\bar{\sigma}}$ sont nuls². La relation (12) devient alors :

$$(13) \quad f(x_0 + \varepsilon a) - f(x_0) = -\lambda_\sigma g^\sigma(x_0 + \varepsilon a) + \frac{1}{2} \varepsilon^2 (a.\nabla)^2 \mathcal{L}(x_0) + \varepsilon^2 o(\varepsilon)$$

Quelle est la condition suffisante pour qu'au voisinage de x_0 nous ayons effectivement :

$$(14) \quad f(x_0 + \varepsilon a) - f(x_0) \leq 0 \quad ?$$

Il apparaît immédiatement, compte tenu du fait que $-\lambda_\sigma g^\sigma(x_0 + \varepsilon a)$ est une quantité négative du premier ordre en ε , ou nulle, que le problème des conditions suffisantes ne se pose que pour les points qui sont sur la frontière du domaine admissible³, c'est à dire tels que :

$$(15) \quad g^\sigma(x_0 + \varepsilon a) = 0 \quad \text{pour tous les } \sigma.$$

La condition suffisante d'optimalité est donc :

$$(16) \quad (a.\nabla)^2 \mathcal{L}(x_0) < 0$$

On peut noter ici encore que la métrique mise en oeuvre ne joue qu'un rôle intermédiaire. L'expression $a.\nabla$ représente en effet dans les diverses formules un opérateur différentiel scalaire, $a^i \partial_i$ en notation tensorielle, qui reste évidemment valable dans un espace simplement affine, comme le sont en général ceux utilisés par les économistes.

G.Ringeisen

octobre 2005

2. Il est intéressant de noter que d'après la relation (8) la nullité d'un $\lambda_{\bar{\sigma}}$ signifie que le vecteur réciproque $a_{\bar{\sigma}}$ associé est orthogonal à $\nabla f(x_0)$, c'est à dire que la surface $f(x)$ est tangente à la ligne de coordonnées curvilignes susceptible d'être associée à $\lambda_{\bar{\sigma}}$.

3. On retrouve avec la relation (13) le cas particulier intéressant où les contraintes saturées sont en nombre n . Les $g^s(x)$ ne peuvent alors être simultanément nuls en-dehors de x_0 , la relation (13) est du premier ordre en ε , et les conditions K-T sont donc nécessaires et suffisantes.