CALCUL DES VARIATIONS (A11)
(24 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **calcul des variations** regroupe des méthodes d'**optimisation** de nature fonctionnelle.

(i) Le calcul des variations met en oeuvre un intervalle I = [a,b] $\subset$ **R**, un **espace normé** E, son **espace affine** associé A, et une fonction continue :

(1) $f : I \times U \mapsto \mathbf{R}$, ie $(t, (m, x)) \mapsto f(t, m, x)$,

où U est un ouvert de A x E (cf **application continue**, **espace vectoriel**, **espace vectoriel normé**, **topologie**).

Si $\varphi : t \mapsto m = \varphi(t)$ est une fonction $I \mapsto A$ de classe C^1 tq (cf **différentiabilité**) :

(2) $(\varphi(t), \varphi'(t)) \in U, \quad \forall t \in I$,

alors, la fonction composée $t \mapsto f(t, \varphi(t), \varphi'(t))$ de I dans **R** est continue. On peut l'intégrer pr à la **mesure de LEBESGUE** λ_1 (avec $d\lambda_1(t) = dt$) :

(3) $J(\varphi) = \int \mathbf{1}_I(t) f(t, \varphi(t), \varphi'(t))\, dt$,

ce qui définit une **fonctionnelle** :

(4) $J : \varphi \mapsto J(\varphi)$ de A^I dans **R**.

Le problème de base du **calcul des variations** est le suivant. Etant donné a, b, f et a = φ (a), b = φ (b), on définit l'ensemble :

(5) $B = \{\varphi \in C^1(I, A) : \varphi(a) = a, \varphi(b) = b\}$,

où $C^1(I, A)$ désigne l'ensemble des fonctions de classe C^1 de I dans A. Il s'agit alors de résoudre le **problème d'optimisation fonctionnelle** :

(6) $\text{opt}_{\varphi \in B} J(\varphi)$

(avec opt = inf, ou opt = sup).

Sous des conditions générales portant sur f et φ (eg fonctions de classe C^2), on montre que, pour que J soit stationnaire (ie minimum ou maximum) au point $\varphi_0 \in B$, il est nécessaire que φ_0 soit solution de l'**équation différentielle** du second ordre, dite **équation de L. EULER - J.L. LAGRANGE** :

(7) $\partial f / \partial m = (d / dt)(\partial f / \partial x)$,

soit, de façon explicite :

(7)' $D_2 f(t, \varphi_0(t), \varphi_0'(t)) = (d/dt)(D_3 f(t, \varphi_0(t), \varphi_0'(t))$,

où D_2 (resp D_3) désigne la dérivée partielle pr au deuxième (resp troisième) argument.

Une telle solution φ_0 est appelée **extrémale** de J.

Diverses extensions ont été établies.

(ii) Cas d'un ***extremum sous contrainte***. Soit $g_1, \ldots, g_k$ des fonctions de même type que f, et $(b_i)_{i=1,\ldots,k}$ des réels donnés. On définit les fonctions suivantes :

(8) $\varphi \mapsto G_i(f) = \int 1_I(t)\, g_i(t, \varphi(t), \varphi'(t))\, dt, \ \forall i \in N_k^*$,

et l'on suppose que :

(a) les dérivées $D G_i$ sont (linéairement) indépendantes au point φ_0 ;

(b) les équations $G_i(\varphi) = b_i$ sont vérifiées au point φ_0 .

Les k équations (8) s'écrivent, sous forme vectorielle :

(8)' $G(\varphi) = \int_I g(t, \varphi(t), \varphi'(t))\, dt = b$.

Alors, pour que J soit extrémale au point $\varphi_0 \in B$, sous les contraintes (8)', il est nécessaire qu'il existe un vecteur $\lambda \in \mathbf{R}^k$ tq l'**équation de L. EULER - J.L. LAGRANGE généralisée** suivante :

(9) $\partial f / \partial m = (d/dt)(\partial f / \partial x) = \sum_{i=1}^{k} \lambda_i \{(\partial g_i / \partial m) - (d/dt)(\partial g_i / \partial x)\}$

soit vérifiée.

Le vecteur λ est appelé **vecteur des multiplicateurs de J.L. LAGRANGE**, ou simplement **multiplicateur (vectoriel) de J.L. LAGRANGE**.

(iii) Cas d'***extrémités variables***. On suppose que a, b, α et β sont variables. Autrement dit, étant donné un intervalle compact $K \subset \mathbf{R}$ et un **espace affine** normé réel F, on définit les applications :

(10)
$$a : F \mapsto K, \quad b : F \mapsto K,$$
$$\alpha : F \mapsto A, \quad \beta : F \mapsto A.$$

Alors, pour que φ_0 et θ_0 , satisfaisant les conditions :

(11)
$$\varphi_0(a(\theta_0)) = \alpha(\theta_0),$$
$$\varphi_0(b(\theta_0)) = \beta(\theta_0),$$

rendent stationnaire la fonctionnelle :

(12) $(\varphi\ ,\ \theta)\ \mapsto\ J\ (\varphi,\ a\ (\theta),\ b\ (\theta))\ =\ \int \mathbf{1}_{[a\ (\theta)\ ,\ b\ (\theta)]}\ f\ (t,\ \varphi\ (t),\ \varphi'\ (t))\ dt$,

il est nécessaire que φ_0 vérifie l'équation de EULER-LAGRANGE (7) et que s_0 vérifie les **conditions de transversalité** suivantes :

$$(13)\quad \begin{cases} (\partial f\ /\ \partial m)(a_0\ ,\ \varphi_0(a_0),\ \varphi_0'(a_0))\ .\ (\partial\alpha - \varphi'(a_0)\ .\ \partial a) + f(a_0\ ,\ \varphi_0(a_0),\ \varphi_0'(a_0))\ .\ \partial a\ =\ 0, \\ (\partial f\ /\ \partial m)(b_0\ ,\ \varphi_0(b_0),\ \varphi_0'(b_0))\ .\ (\partial\beta - \varphi'(b_0)\ .\ \partial b) + f(b_0\ ,\ \varphi_0(b_0),\ \varphi_0'(b_0))\ .\ \partial b\ =\ 0, \end{cases}$$

dans lesquelles $a_0 = a(\theta_0)$, $b_0 = b(\theta_0)$, et où ∂a, ∂b, $\partial\alpha$ et $\partial\beta$ sont les différentielles respectives des fonctions a, b, α et β lorsque θ varie de $\partial\theta$ pr à θ_0 .

(iv) Ce qui précède s'étend au cas où t prend ses valeurs dans un ouvert de $\mathbf{R}^p$ (au lieu de $I \subset \mathbf{R}$). Ceci conduit à des **équations aux dérivées partielles** du second ordre.

Il s'étend aussi au cas où f est de la forme :

(14) $f : (t,\ (m,\ x_1\ ,...,\ x_p))\ \mapsto\ f\ (t,\ m,\ x_1\ ,...,\ x_p)$,

avec $x_i = \varphi^{(j)}\ (t)\ (\forall\ j \in N_p{}^*)$, ie dépend des p dérivées successives de φ, supposée de classe C^p. Ce dernier cas, d'apparence plus générale, se réduit au cas décrit plus haut en définissant, comme dans l'étude des **équations différentielles** d'ordre p, les **fonctions auxiliaires** (qui jouent un rôle d'inconnues nouvelles) :

$$(15)\quad \begin{cases} \varphi_1 = \varphi\ ', \\ \quad ... \\ \varphi_p = \varphi^{(p)}. \end{cases}$$

(iv) En **Statistique**, les **méthodes variationnelles** interviennent dans divers contextes (eg en **statistique non paramétrique**), dans lesquels l' « inconnue » du problème d'optimisation est souvent une **densité de probabilité** (ie sa **classe d'équivalence**) ou une **statistique**. Souvent même, la dérivée n'intervient pas explicitement dans l'intégrande. Ainsi, dans le cas où l'on doit résoudre le problème :

(16) $\max_f\ J\ (\phi)\ =\ \int f\ (\phi\ (t))\ dt \qquad \text{sous}\quad K\ (\phi)\ =\ \int g\ (\phi\ (t))\ dt\ =\ c$,

dans lequel f est une densité de probabilité (pr à la mesure de LEBESGUE dt de $\mathbf{R}$) (inconnue), $\phi : \mathbf{R} \mapsto \mathbf{R}$ une fonction donnée tq f o ϕ est dt-intégrable, J la fonctionnelle correspondante, à valeurs dans $\mathbf{R}$, $g : \mathbf{R}_+ \mapsto \mathbf{R}^m$ une fonction vectorielle donnée, K la fonctionnelle correspondante et $c \in \mathbf{R}^m$ une constante donnée. Le **lagrangien** s'écrit :

(17) $L\ (\phi,\ \lambda)\ =\ J\ (\phi) - \lambda'\ K\ (\phi)$, $\qquad$ avec $\lambda \in \mathbf{R}^m$.

On considère alors une fonction ψ : [a, b] $\mapsto$ **R** tq $\int \psi$ (t) dt = 0 et l'on pose :

(18) $G(a) = L(\phi + a\,\psi, \lambda)$.

Si J atteint un maximum en $\phi = \phi^*$, alors G (a) atteint un maximum pour a = 0. Par suite, la dérivée G' (a) doit vérifier la condition du premier ordre pour un extremum (ie ici G' (0) = 0). On cherche ensuite à écrire G' (0) sous la forme :

(19) $\int L(\phi)\, g(t)\, dt = 0$.

Comme cette forme est vérifiée pour toute fonction g (d'intégrale nulle), il en résulte que L (ϕ) = g (pour une certaine constante g), d'où la solution $\phi^* = L^{-1}(g)$. On vérifie ensuite que f* minimise bien J (ϕ), ie que G" (0) $<$ 0 (condition du second ordre pour un maximum de G).

A titre d'exemples, si ξ est une **vars** et f sa **densité de probabilité** :

(a) maximiser l'**entropie de SHANNON** (cf **information de SHANNON**) :

(20) $J(f) = -\int \mathbf{1}_{[a,\, b]}(t) \,.\, \{t \,.\, f(t)\} \,.\, \mathrm{Log}\, \{t \,.\, f(t)\}\, dt$

sous les contraintes eg E $\xi = \mu$ et V $\xi = \sigma^2$ (donnés) conduit à une solution f* qui n'est autre que la densité de la **loi normale** $\mathcal{N}_1(\mu, \sigma^2)$;

(b) maximiser cette même entropie (20) sous les contraintes E $\xi = \mu$ et $\int g(t)\, f(t)\, dt = \nu$ (avec g : [a, b] $\mapsto$ **R**, où μ et ν sont donnés) conduit à une solution de la forme « hyper-exponentielle » :

(21) $f^*(t) = \mathbf{1}_{[a,b]}(t) \,.\, (c / t) \,.\, \exp \{-(\beta / t) \,.\, (h(t) - \alpha)\}, \qquad \forall\, t \in \mathbf{R}$,

dans laquelle c $>$ 0 est une **constante** de **normalisation** (qui dépend en général de h, α et β).