

Calcul différentiel et optimisation

Dans tout le chapitre, E et F désignent des \mathbb{R} -espaces vectoriels normés de dimension finie nulle, \mathcal{U} désigne un ouvert de E .

1 DIFFÉRENTIELLE

1 Différentielle en un point

Rappel : si I est intervalle de \mathbb{R} , $f : I \rightarrow F$ est dérivable en $a \in I$ si et seulement si f admet un développement limité à l'ordre 1 en a si et seulement si on a un vecteur $b \in F$ tel que

$$f(a+h) = f(a) + hb + o_{h \rightarrow 0_{\mathbb{R}}}(h)$$

(et, dans ce cas, $b = f'(a)$).

Nous allons généraliser cette idée aux fonctions définies sur un espace vectoriel normé de dimension finie E .

Définition 1 : Application différentiable en un point

Soit f définie sur un ouvert \mathcal{U} de E , à valeurs dans F . Soit a un point de \mathcal{U} . On dit que f est **différentiable** en a lorsqu'il existe une application linéaire l_a de E dans F telle que, au voisinage de 0_E ,

$$f(a+h) = f(a) + l_a(h) + o_{h \rightarrow 0_E}(\|h\|_E) \leftarrow \begin{matrix} \|h\|_E \varepsilon(h) \\ \text{où } \varepsilon(h) \xrightarrow{h \rightarrow 0_E} 0_F \end{matrix}$$

ou encore, au voisinage de a ,

$$f(x) = f(a) + l_a(x-a) + o_{x \rightarrow a}(\|x-a\|_E)$$

Lorsqu'elle existe, l'application l_a est unique et appelée **différentielle** de f au point a ou encore **application linéaire tangente** à f en a , notée $df(a) \in \mathcal{L}(E, F)$. On a donc

$$f(a+h) = f(a) + df(a)(h) + o_{h \rightarrow 0_E}(\|h\|_E)$$

Remarque

R1 – Lorsque cela a du sens, on définit donc une application $df : \begin{matrix} E & \rightarrow & \mathcal{L}(E, F) \\ a & \rightarrow & df(a) \end{matrix}$ appelée **différentielle** de f .

R2 – On note parfois $df(a) \cdot h$ au lieu de $df(a)(h)$.

R3 – L'unicité vient du

$$df_a(h) = df(a)(h)$$

Lemme 1

Si $\varphi \in \mathcal{L}(E, F)$ est telle que $\varphi(h) = o_{h \rightarrow 0}(h)$, alors φ est l'application nulle.

Démonstration

$$\varphi(h) = \|h\| \varepsilon(h) \text{ avec } \varepsilon(h) \xrightarrow{h \rightarrow 0_E} 0_F.$$

Alors pour tout $x \in E$, $t \in \mathbb{R}_+^*$, $t\varphi(x) = \varphi(tx) = t\|x\|\varepsilon(tx)$ donc $\varphi(x) = \|x\| \varepsilon(tx) \xrightarrow{t \rightarrow 0} 0_F$.

Donc pour tout $x \in E$, $\varphi(x) = 0_F$. ■

Propriété 1 : différentiable \Rightarrow continue

Si f est différentiable en a , alors f est continue en a .

Démonstration

Par continuité de l'application linéaire en dimension finie (celle de E suffit) $df(a)$, en passant à la limite dans le développement limité, $f(a+h) - f(a) \xrightarrow{h \rightarrow 0_E} 0_F$. ■

2 Cas particuliers

Propriété 2 : Cas d'une fonction d'une variable réelle

Dans le cas d'une fonction $f : I \rightarrow F$, f est dérivable en a si et seulement si elle est différentiable en a .

Dans ce cas,

$$df(a) : h \mapsto f'(a)h$$

$$f(a+h) = f(a) + f'(a)h + o_{h \rightarrow 0}(h) \text{ où } h \mapsto f'(a)h \text{ linéaire}$$

**Propriété 3 : Cas d'une fonction constante**

Si $f : \mathcal{U} \subset E \rightarrow F$ est constante, elle est différentiable en tout point de \mathcal{U} et pour tout $a \in \mathcal{U}$,

$$df(a) = 0_{\mathcal{L}(E, F)}$$

Démonstration

$f(a+h) = f(a) + 0_F + 0_F$ avec $h \mapsto 0_F$ linéaire et $0_F = o(h)$.

Donc f est différentiable en a et $df(a) = 0_{\mathcal{L}(E, F)}$. ■

Propriété 4 : Cas d'une fonction linéaire

Si $f : \mathcal{U} \subset E \rightarrow F$ est la restriction à \mathcal{U} d'une application linéaire $\varphi \in \mathcal{L}(E, F)$, elle est différentiable en tout point de \mathcal{U} et pour tout $a \in \mathcal{U}$,

$$df(a) = \varphi$$

Démonstration

$f(a+h) = \varphi(a+h) = \varphi(a) + \varphi(h) + 0_F$ avec φ linéaire et $0_F = o(h)$.

Donc f est différentiable en a et $df(a) = \varphi$ (donc df est constante.) ■

Exercice 1 : Montrer que $f : \begin{matrix} \mathcal{M}_n(\mathbb{R}) & \longrightarrow & \mathcal{M}_n(\mathbb{R}) \\ M & \longmapsto & M^2 \end{matrix}$ est différentiable en toute $A \in \mathcal{M}_n(\mathbb{R})$ et calculer $df(A)$.

Exercice 2 : Montrer que, si E est un espace euclidien $f : \begin{matrix} E & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \|x\|^2 \end{matrix}$ est différentiable en toute $a \in E$ et calculer $df(a)$.

Exercice 3 : Montrer que, si E est un espace euclidien, $u \in \mathcal{L}(E)$, $f : \begin{matrix} E & \longrightarrow & \mathbb{R} \\ x & \longmapsto & (x|u(x)) \end{matrix}$ est différentiable en toute $a \in E$ et calculer $df(a)$. Que se passe-t-il si, de plus, u est symétrique ?

3 Dérivée selon un vecteur, dérivées partielles

Le fait de travailler sur un ouvert \mathcal{U} assure, pour a point de \mathcal{U} et v vecteur de E , l'existence d'un $\delta > 0$ (« distance de sécurité ») tel que pour tout $t \in]-\delta, \delta[$, $a+tv \in \mathcal{U}$ (on s'éloigne de a dans la direction de v), cela permet de définir l'application $\phi : t \mapsto f(a+tv)$ d'une variable réelle au voisinage de 0.

Définition 2 : Dérivée selon un vecteur

On dit que $f : \mathcal{U} \subset E \rightarrow F$ est **dérivable selon le vecteur** $v \in E$ au point $a \in \mathcal{U}$, lorsque

$\phi : t \mapsto f(a+tv)$ dérivable en 0.

On note alors

$$D_v f(a) = \phi'(0) = \lim_{t \rightarrow 0} \frac{f(a+tv) - f(a)}{t}$$

Exemple

E1 – La fonction $f : (x, y) \in \mathbb{R}^2 \mapsto \frac{xy^2}{x^2+y^4}$ si $(x, y) \neq (0, 0)$ et 0 sinon admet des dérivées selon tout vecteur en $(0, 0)$ et si $v = (\alpha, \beta) : D_v f(0, 0) =$

Définition 3 : Dérivées partielles

Soient $\mathcal{B} = (e_1, \dots, e_p)$ une base de E , $f : \mathcal{U} \subset E \rightarrow F$, $a \in \mathcal{U}$, $j \in \llbracket 1, p \rrbracket$.

On appelle **j° dérivée partielle de f en a** , lorsqu'elle existe, la dérivée de f selon le vecteur e_j de base en a :

$$\partial_j f(a) = \frac{\partial f}{\partial x_j}(a) = D_{e_j} f(a) \in F.$$

Remarque

R4 – Les dérivées partielles vues pour les fonctions définies sur un ouvert de \mathbb{R}^p sont les dérivées partielles dans la base canonique de \mathbb{R}^p .

Bien sûr la notion générale de dérivée partielle dépend de la base choisie. En ce sens, la notation $D_{e_j} f(a)$ est la plus précise.

R5 – Comme on en a déjà l'habitude, calculer la j° dérivée partielle revient à fixer les coordonnées selon tous les autres vecteurs de bases et dériver par rapport à la seule variable x_j . : en effet, en dérive

$$\phi : t \mapsto f(a + te_j) = f(a_1 e_1 + \dots + (a_j + t)e_j + \dots + a_n e_n)$$

en 0, ce qui revient aussi à dériver

$$t \mapsto f(a_1 e_1 + \dots + t e_j + \dots + a_n e_n)$$

en a_j .

Exemple

E2 – Avec $f : (x, y) \in \mathbb{R}^2 \mapsto \frac{xy}{x^2 + y^2}$ si $(x, y) \neq (0, 0)$ et $\frac{0}{0}$ sinon, on voit que Δ on peut avoir des dérivées partielles en $(0, 0)$ sans avoir de dérivée selon certains vecteurs ($v = (\alpha, \beta) \in \mathbb{R}_*^2$, ici).

4 Lien entre différentielle et dérivées partielles

Propriété 5 : Lien entre différentielle et dérivées partielles

Soit $f : \mathcal{U} \rightarrow F$, $a \in \mathcal{U}$. Si f est différentiable en a , alors f est dérivable en a selon tout vecteur $v \in E$ et $D_v f(a) = df(a)(v)$

Démonstration

$$f(a + tv) = f(a) + df(a)(tv) + o_{t \rightarrow 0}(t) \text{ donc } \frac{f(a + tv) - f(a)}{t} \xrightarrow{t \rightarrow 0} df(a)(v) \text{ par linéarité de } df(a).$$

Cas particulier 1 : Dérivée selon un vecteur de base

Soit $f : \mathcal{U} \rightarrow F$, $a \in \mathcal{U}$, $\mathcal{B} = (e_1, \dots, e_p)$ une base de E . On note $\frac{\partial f}{\partial x_j}$ les dérivées partielles de f dans la base \mathcal{B} . Si f est différentiable en a , alors pour tout $j \in \llbracket 1, p \rrbracket$,

$$df(a)(e_j) = \frac{\partial f}{\partial x_j}(a)$$

Propriété 6 : Expression de la différentielle avec les dérivées partielles

Soit $\mathcal{B} = (e_1, \dots, e_p)$ une base de E , $f : \mathcal{U} \rightarrow F$, $a \in \mathcal{U}$ tel que f est différentiable en a .

Alors pour tout vecteur $h = \sum_{j=1}^p h_j e_j \in E$,

$$df(a)(h) = \sum_{j=1}^p h_j \frac{\partial f}{\partial x_j}(a)$$

On note $dx_j : \begin{cases} E & \rightarrow \mathbb{R} \\ h & \mapsto h_j \end{cases}$ la forme linéaire j^{e} coordonnée dans \mathcal{B} . Alors on a

$$df(a) = \sum_{j=1}^p \frac{\partial f}{\partial x_j}(a) dx_j$$

Démonstration

C'est la linéarité de $df(a)$:

$$df(a)\left(\sum_{j=1}^p h_j e_j\right) = \sum_{j=1}^p h_j df(a)(e_j) = \sum_{j=1}^p h_j \frac{\partial f}{\partial x_j}(a).$$

Remarque

R6 – Le DL₁ de f différentiable en a s'écrit alors

$$f(a + h) = f(a) + \sum_{j=1}^p h_j \frac{\partial f}{\partial x_j}(a) + o(h).$$



5 Matrice jacobienne

Définition 4 : Matrice jacobienne

Soit $p = \dim E$, $n = \dim F$, \mathcal{U} ouvert de E et $f : E \rightarrow F$ différentiable, $\mathcal{B} = (e_1, \dots, e_p)$ une base de E et $\mathcal{C} = (\varepsilon_1, \dots, \varepsilon_n)$ une base de F . On note $f = \sum_{i=1}^n f_i \varepsilon_i$.

On appelle **matrice jacobienne** de f en a dans les bases \mathcal{B} et \mathcal{C} la matrice

$$J_f(a) = \left(\frac{\partial f_i}{\partial x_j}(a) \right)_{(i,j) \in \llbracket 1, n \rrbracket \times \llbracket 1, p \rrbracket}.$$

Propriété 7 : Matrice jacobienne et différentielle

$$J_f(a) = \text{Mat}_{\mathcal{B}, \mathcal{C}}(df(a))$$

Démonstration

$$df(a)(e_j) = \frac{\partial f}{\partial x_j}(a) = \sum_{i=1}^p \frac{\partial f_i}{\partial x_j}(a) \varepsilon_i$$

permet bien de remplir la j^{e} colonne de $J_f(a)$. ■

6 Gradient

Définition 5 : Gradient

Soit E un espace **euclidien**, \mathcal{U} un ouvert de E , $f : \mathcal{U} \rightarrow \mathbb{R}$ une fonction différentiable. Alors pour tout $a \in \mathcal{U}$, il existe un unique vecteur noté $\nabla f(a)$ ou $\overrightarrow{\text{grad}} f(a)$ et appelé **gradient de f en a** tel que pour tout $h \in E$,

$$df(a)(h) = (\nabla f(a)|h).$$

Démonstration

C'est le théorème de représentation de Riesz : la **forme** linéaire $df(a)$ s'écrit $x \mapsto (b|x)$ pour un certain vecteur b ... ■

Remarque

R7 – On comprend mieux la notation alternative $df(a) \cdot h$ faisant penser à un produit scalaire. Le DL_1 de f différentiable en a dans E euclidien se récrit

$$f(a+h) = f(a) + \underbrace{(\nabla f(a)|h)}_{\text{produit scalaire}} + o(h)$$

R8 – Dans \mathbb{R}^2 muni du produit scalaire canonique,

$$df(a)(h) = h_1 \frac{\partial f}{\partial x}(a) + h_2 \frac{\partial f}{\partial y}(a) = \begin{pmatrix} \frac{\partial f}{\partial x}(a) \\ \frac{\partial f}{\partial y}(a) \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

On retrouve notre gradient habituel.

Exactement le même raisonnement (expression du produit scalaire en base orthonormée) donne la propriété suivante.

Propriété 8 : Coordonnées du gradient

Soit E un espace **euclidien**, \mathcal{U} un ouvert de E , $f : \mathcal{U} \rightarrow \mathbb{R}$ une fonction différentiable. Si l'on fixe une base **orthonormée** de E , les coordonnées de $\nabla f(a)$ dans cette base sont

$$\left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_p}(a) \right).$$

Remarque : Interprétation géométrique

R9 – On suppose f différentiable en a et $\nabla f(a) \neq 0_E$.

Notons S la sphère unité de E pour la norme euclidienne $\|\cdot\|$. D'après l'inégalité de Cauchy-Schwarz

$$\forall h \in S, df(a)(h) = (\nabla f(a)|h) \leq |(\nabla f(a)|h)| \leq \|\nabla f(a)\| \|h\| = \|\nabla f(a)\|.$$

Or ce majorant est atteint pour $h_0 = \frac{\nabla f(a)}{\|\nabla f(a)\|}$.

Ainsi le vecteur $\nabla f(a)$ donne la direction et le sens de plus forte variation de la fonction f .

C'est à la base d'algorithmes d'optimisation (descente de gradient), utilisés par exemple en Machine Learning (apprentissage automatique).

Ci-après, on représente, au cœur de Mafate, des « lignes de niveau » de la fonction $(x, y) \mapsto f(x, y)$ qui au point de coordonnées (x, y) associe son altitude (on parle d'isoplèthes d'altitude).

Ces lignes de niveau sont des courbes d'équation $f(x, y) = \text{constante}$.

Tracer la direction et le sens de ∇f en quelques points.

On peut montrer que le gradient est normal aux lignes de niveau en tous points.



II OPÉRATIONS SUR LES DIFFÉRENTIELLES

1 Combinaisons linéaires

Propriété 9 : Linéarité

Soit \mathcal{U} ouvert de E , $f, g : \mathcal{U} \rightarrow F$ différentiables en $a \in \mathcal{U}$, $\lambda, \mu \in \mathbb{R}$.
Alors $\lambda f + \mu g$ est différentiable et

$$d(\lambda f + \mu g) = \lambda df + \mu dg$$

Démonstration

Il suffit de remarquer, par propriétés de la dérivation que $J_{\lambda f + \mu g}(a) = \lambda J_f(a) + \mu J_g(a)$ et de repasser aux applications linéaires.

Autre possibilité : on ajoute les DL₁ :

$$\lambda f(a+h) + \mu g(a+h) = \lambda f(a) + \mu g(a) + \lambda df(a)(h) + \mu dg(a)(h) + o(h)$$

avec $\lambda df(a) + \mu dg(a)$ linéaire, donc par unicité, $\lambda f + \mu g$ est différentiable et

$$d(\lambda f + \mu g)(a) = \lambda df(a) + \mu dg(a).$$

Remarque

R 10 - $f \mapsto df$ est donc linéaire.

2 Image par une application multilinéaire

Propriété 10 : Image par une application multilinéaire

Soit \mathcal{U} ouvert d'un \mathbb{R} -espace vectoriel normé de dimension finie, E_1, \dots, E_q, F des \mathbb{R} -espaces vectoriels normés de dimension finie, $f_1 : \mathcal{U} \rightarrow E_1, \dots, f_q : \mathcal{U} \rightarrow E_q$ des applications différentiables en $a \in \mathcal{U}$ et $M : E_1 \times \dots \times E_q \rightarrow F$ une fonction q -linéaire.

Alors $\phi : \mathcal{U} \rightarrow F$ définie par $\phi : x \mapsto M(f_1(x), \dots, f_q(x))$ est différentiable en a et

$$d\phi(a) : h \mapsto \sum_{k=1}^q M(f_1(a), \dots, df_k(a)(h), \dots, f_q(a))$$

Démonstration

On traite le cas bilinéaire. Le cas général est analogue.

1^{re} tentative On peut voir le faire avec les dérivées selon les vecteurs. On sait, sous réserve d'existence, que $d\phi(a)(h) = D_h\phi(a)$. Or

$$\psi : t \mapsto \phi(a+th) = B(f(a+th), g(a+th))$$

est dérivable en 0 (car f et g sont différentiables en a) de dérivée

$$\begin{aligned} \psi'(0) &= D_h\phi(a) = B(D_h f(a), g(a)) + B(f(a), D_h g(a)) \\ &= B(df(a)(h), g(a)) + B(f(a), dg(a)(h)) \end{aligned}$$

Le problème est que l'existence de dérivée selon tout vecteur ne garantit pas la différentiabilité...

On peut par exemple montrer que la fonction $f : (x, y) \mapsto \begin{cases} \frac{y^2}{x} & \text{si } x \neq 0 \\ y & \text{sinon} \end{cases}$ admet des

dérivées selon tout vecteur en $(0,0)$ mais n'est pas continue, et encore moins différentiable.

Bonne méthode On utilise les DL₁ : avec $\varepsilon_1(h) \xrightarrow{h \rightarrow 0} 0$ et $\varepsilon_2(h) \xrightarrow{h \rightarrow 0} 0$,

$$\begin{aligned} \phi(a+h) &= B(f(a+h), g(a+h)) \\ &= B(f(a) + df(a)(h) + \|h\| \varepsilon_1(h), g(a) + dg(a)(h) + \|h\| \varepsilon_2(h)) \\ &= B(f(a), g(a)) + B(df(a)(h), g(a)) + B(f(a), dg(a)(h)) + \psi(h) \end{aligned}$$

avec $h \mapsto B(df(a)(h), g(a)) + B(f(a), dg(a)(h))$ linéaire et

$$\begin{aligned} \psi(h) &= \|h\| B(f(a) + df(a)(h), \varepsilon_2(h)) + \|h\| B(\varepsilon_1(h), g(a) + dg(a)(h)) \\ &\quad + \|h\|^2 B(\varepsilon_1(h), \varepsilon_2(h)) + B(df(a)(h), dg(a)(h)) \end{aligned}$$



donc

$$\frac{\psi(h)}{\|h\|} = B(f(a) + df(a)(h), \varepsilon_2(h)) + B(\varepsilon_1(h), g(a) + dg(a)(h)) + \|h\| B(\varepsilon_1(h), \varepsilon_2(h)) + \frac{B(df(a)(h), dg(a)(h))}{\|h\|}.$$

Les trois premiers termes tendent vers 0 lorsque $h \rightarrow 0$ par continuité de l'application bilinéaire en dimension finie.

Pour le dernier, on remarque que

$$(x, y) \mapsto B(df(a)(x), dg(a)(y))$$

est bilinéaire, donc on a, par continuité, un réel C tel que

$$\left\| \frac{B(df(a)(h), dg(a)(h))}{\|h\|} \right\| = \frac{\|B(df(a)(h), dg(a)(h))\|}{\|h\|} \leq \frac{C \|h\|^2}{\|h\|} = C \|h\| \xrightarrow{h \rightarrow 0} 0.$$

Donc $\psi(h) = o_{h \rightarrow 0}(h)$, ce qui permet d'obtenir toute la conclusion. ■

3 Composition

Propriété 11 : Différentielle d'une composée

Soit \mathcal{U} ouvert de E , \mathcal{V} ouvert de F , $f: \mathcal{U} \rightarrow \mathcal{V}$ différentiable en $a \in \mathcal{U}$ et $g: \mathcal{V} \rightarrow G$ différentiable en $b = f(a)$.

Alors $g \circ f$ est différentiable en a et

$$d(g \circ f)(a) = dg|_{f(a)} \circ df(a)$$

Propriété 12 : Matrice jacobienne d'une composée

En munissant E , F et G de bases, on obtient

$$J_{g \circ f}(a) = J_g(f(a)) \times J_f(a)$$

Remarque

R 11 – D'où la règle de la chaîne.

Démonstration

Il suffit de composer les DL₁ : on écrit, avec $\varepsilon_1(h) \xrightarrow{h \rightarrow 0} 0$ et $\varepsilon_2(k) \xrightarrow{k \rightarrow 0} 0$, et pour simplifier la lecture, les applications linéaires $\varphi = df(a)$ et $\psi = dg(f(a))$,

$$f(a+h) = f(a) + \varphi(h) + \|h\| \varepsilon_1(h)$$

et

$$g(f(a) + k) = g(f(a)) + \psi(k) + \|k\| \varepsilon_2(k)$$

puis

$$\begin{aligned} g(f(a+h)) &= g\left[f(a) + \underbrace{(\varphi(h) + \|h\| \varepsilon_1(h))}_k\right] \\ &= g(f(a)) + \psi(\varphi(h) + \|h\| \varepsilon_1(h)) \\ &\quad + \|\varphi(h) + \|h\| \varepsilon_1(h)\| \varepsilon_2(\varphi(h) + \|h\| \varepsilon_1(h)) \\ &= g \circ f(a) + [\psi \circ \varphi](h) + \zeta(h) \end{aligned}$$

avec

$$\frac{\zeta(h)}{\|h\|} = \psi(\varepsilon_1(h)) + \frac{\|\varphi(h) + \|h\| \varepsilon_1(h)\|}{\|h\|} \varepsilon_2(\varphi(h) + \|h\| \varepsilon_1(h)).$$

φ et ψ étant linéaires en dimension finie, elles sont continues, donc $\psi(\varepsilon_1(h)) \xrightarrow{h \rightarrow 0} 0$,

$\varphi(h) \xrightarrow{h \rightarrow 0} 0$ puis $\varepsilon_2(\varphi(h) + \|h\| \varepsilon_1(h)) \xrightarrow{h \rightarrow 0} 0$.

Enfin,

$$\frac{\|\varphi(h) + \|h\| \varepsilon_1(h)\|}{\|h\|} \leq \frac{\|\varphi(h)\|}{\|h\|} + \varepsilon_1(h)$$

et par continuité de l'application linéaire φ , on a $C \in \mathbb{R}$ tel que pour tout h , $\|\varphi(h)\| \leq C \|h\|$

donc $\frac{\|\varphi(h) + \|h\| \varepsilon_1(h)\|}{\|h\|}$ est borné et finalement $\frac{\zeta(h)}{\|h\|} \xrightarrow{h \rightarrow 0} 0$, c'est-à-dire $\zeta(h) = o(h)$.

Par unicité, on en déduit que $g \circ f$ est différentiable en a et

$$d(g \circ f)(a) = \psi \circ \varphi = dg(f(a)) \circ df(a). \quad \blacksquare$$

Corollaire 1 : Dérivée le long d'un arc

Soit \mathcal{U} ouvert de E , $f: \mathcal{U} \rightarrow F$, I intervalle de \mathbb{R} et $\gamma: I \rightarrow \mathcal{U}$ une application dérivable.

On suppose que f est différentiable en $\gamma(t)$ où $t \in I$. Alors $f \circ \gamma$ est dérivable en t et

$$(f \circ \gamma)'(t) = df|_{\gamma(t)}(\gamma'(t))$$

$$\text{the } E, d(g \circ f)(a)(h) = dg|_{f(a)}(df(a)(h))$$

En munissant E d'une base, on retrouve l'expression vue avec la règle de la chaîne, si les coordonnées de $\gamma(t)$ sont $(\gamma_1(t), \dots, \gamma_p(t))$,

$$(f \circ \gamma)'(t) = \sum_{i=1}^p \gamma_i'(t) \frac{\partial b}{\partial x_i}(\gamma(t))$$

En particulier, si $\gamma: t \mapsto x + tv$ où $x, v \in \mathcal{U}$ sont fixés,

$$(f \circ \gamma)'(t) = df(x + tv)(v)$$

III $\phi: t \mapsto f(a + tv) = D_v f(\gamma(t))$ on dérive dans la direction v
CLASSE \mathcal{C}^1

Définition 6 : Fonctions de classe \mathcal{C}^1

$f: \mathcal{U} \rightarrow F$ est dite **de classe \mathcal{C}^1** lorsque f est différentiable sur \mathcal{U} et $df: \mathcal{U} \rightarrow \mathcal{L}(E, F)$ est continue.

Propriété 13 : Opérations

Toute combinaison linéaire, toute composée d'applications de classe \mathcal{C}^1 l'est encore.

Si M est q -linéaire et f_1, \dots, f_q sont \mathcal{C}^1 , $M(f_1, \dots, f_q)$ l'est.

Propriété 14 : Caractérisation par les applications coordonnées

Soit $f: \mathcal{U} \rightarrow F$, d'applications coordonnées (f_1, \dots, f_n) dans une base de F . f est de classe \mathcal{C}^1 si et seulement si toutes les f_i le sont.

Théorème 1 : IMPORTANT - Caractérisation avec les dérivées partielles

f est de classe \mathcal{C}^1 si et seulement si, dans une base quelconque de E , toutes les dérivées partielles de f existent et sont continues.

Démonstration

Non exigible.

Propriété 15 : Expression intégrale le long d'un arc

Soit $f \in \mathcal{C}^1(\mathcal{U}, F)$, $\gamma \in \mathcal{C}^1([0, 1], \mathcal{U})$, $a = \gamma(0)$ et $b = \gamma(1)$. Alors

$$f(b) - f(a) = \int_0^1 df(\gamma(t))(\gamma'(t)) dt$$

Démonstration

On a vu que $d(f \circ \gamma)(\gamma'(t)) = (f \circ \gamma)'(t)$, c'est donc une simple application du théorème fondamental de l'analyse.

Corollaire 2 : Cas particulier $\gamma(t) = a + tv$

En particulier, avec $\gamma(t) = a + tv$ sur $[0, 1]$,

$$f(a + v) - f(a) = \int_0^1 df(a + tv)(v) dt$$

Corollaire 3 : Caractérisation des fonctions constantes sur un ouvert connexe par arcs

Si $f \in \mathcal{C}^1(\mathcal{U}, F)$ et \mathcal{U} est un ouvert **connexe par arcs**, alors

f est constante si et seulement si $df = 0$.

Démonstration

Le sens direct a déjà été vu.
 Pour le sens réciproque, seul le cas convexe est au programme. On suppose donc \mathcal{U} convexe et $df = 0$.

Soient $a, b \in \mathcal{U}$, on veut montrer que $f(a) = f(b)$.

Mais on peut relier a et b par un segment inclus dans \mathcal{U} par convexité : $\gamma: t \in [0, 1] \mapsto (1-t)a + tb$ de classe \mathcal{C}^1 .

Alors, la propriété précédente donne $f(b) - f(a) = 0$.



IV VECTEURS TANGENTS

Définition 7 : Vecteur tangent

Si X est une partie de E , $a \in X$. Un vecteur $v \in E$ est dit **tangent** à X en a lorsqu'il existe $\varepsilon > 0$ et un arc $\gamma :]-\varepsilon, \varepsilon[\rightarrow X$, dérivable en 0 tel que $\gamma(0) = a$ et $\gamma'(0) = v$.
On note $T_a X$ l'ensemble des vecteurs tangents à X en a .

Exemple

E3 – Si \mathcal{G} est un sous-espace affine de E de direction G (sous-espace vectoriel de E), pour tout $a \in \mathcal{G}$,

$$T_a \mathcal{G} = G$$

En effet, on peut écrire $\mathcal{G} = a + G$.

Si $v \in G$, soit $\gamma : \begin{cases}]-1, 1[\rightarrow \mathcal{G} \\ t \rightarrow a + tv \end{cases}$ est dérivable en 0, $\gamma(0) = a$ et $\gamma'(0) = v$.

Réciproquement, si $v \in T_a \mathcal{G}$, on a $\varepsilon > 0$ et $\gamma :]-\varepsilon, \varepsilon[\rightarrow \mathcal{G}$ tel que $\gamma(0) = a$ et $\gamma'(0) = v$.

Alors v est la limite lorsque $t \rightarrow 0$ de $\frac{\gamma(t) - a}{t} \in G$ et comme G est un sous-espace de E qui est de dimension finie, il l'est aussi, donc il est fermé et donc $v \in G$.

E4 – **Sphère d'un espace euclidien** : si $a \in S(0_E, 1)$,

$$T_a S(0_E, 1) = a^\perp$$

En effet, si $v \in a^\perp$, soit $\gamma : \begin{cases}]-1, 1[\rightarrow S(0_E, 1) \\ t \rightarrow \frac{a + tv}{\|a + tv\|} \end{cases}$ est dérivable en 0, $\gamma(0) = a$ et $\gamma'(0) = v$

car, avec $\|a\| = 1$,

$$\gamma : t \mapsto \frac{a + tv}{\sqrt{\|a + tv\|^2}} = \frac{a + tv}{\sqrt{1 + 2t(a|v) + t^2 \|v\|^2}}$$

donc

$$\gamma'(0) = \frac{\|a + 0 \cdot v\| v - \frac{2(a|v) + 2 \cdot 0 \cdot \|v\|^2}{2\|a + 0 \cdot v\|} (a + 0 \cdot v)}{\|a + 0 \cdot v\|^2}$$

Réciproquement, si $v \in T_a S(0_E, 1)$, on a $\varepsilon > 0$ et $\gamma :]-\varepsilon, \varepsilon[\rightarrow S(0_E, 1)$ tel que $\gamma(0) = a$ et $\gamma'(0) = v$.

Alors $\frac{\gamma(t) - a}{t} \xrightarrow[t \rightarrow 0]{} v$ donc

$$\begin{aligned} \left(\frac{\gamma(t) - a}{t} \middle| a \right) &= \frac{(\gamma(t)|a) - 1}{t} = \frac{(\gamma(t)|a) - \gamma(t)|a}{t} \\ &= - \left(\gamma(t) \middle| \frac{\gamma(t) - a}{t} \right) \xrightarrow[t \rightarrow 0]{} (v|a) = -(a|v) \end{aligned}$$

Donc $(v|a) = 0 : v \in a^\perp$.

Comme dernier exemple, on va obtenir que le plan tangent à une surface d'équation $z = f(x, y)$ admet comme vecteur normal le gradient (non nul) de $g : (x, y, z) \mapsto f(x, y) - z$.

Cela revient à transformer l'équation explicite $z = f(x, y)$ en équation implicite $g(x, y, z) = 0$.

Propriété 16 : Plan tangent pour une surface explicite

Soit \mathcal{U} ouvert de \mathbb{R}^2 , $f : \mathcal{U} \rightarrow \mathbb{R}$ différentiable.

Soit $S \subset \mathbb{R}^3$ la surface représentative de f , c'est-à-dire

$$S = \{(x, y, z) \in \mathbb{R}^3, (x, y) \in \mathcal{U} \text{ et } z = f(x, y)\}.$$

$$\text{Soit } g : \begin{cases} \mathcal{U} \times \mathbb{R} \rightarrow \mathbb{R} \\ (x, y, z) \mapsto f(x, y) - z \end{cases}$$

L'ensemble $T_a S$ des vecteurs tangents à S en $a = (x_0, y_0, z_0) \in S$ est un plan vectoriel $P = \nabla g(a)^\perp$ de vecteur normal le vecteur (non nul) $\nabla g(a)$, donc d'équation

$$(\nabla g(a) | (x, y, z)) = 0.$$

On appelle **plan tangent à S en a** le plan affine $a + P$ passant par a et de direction l'ensemble P des vecteurs tangents à X en a , d'équation

$$(\nabla g(a) | (x - x_0, y - y_0, z - z_0)) = 0$$

Remarque

R12 – On retrouve, pour le plan tangent, une équation type « équation de la tangente » :

Démonstration

On raisonne toujours par double inclusion.

- Si $v = (x, y, z) \in T_a S$, on a $\varepsilon > 0$ et $\gamma :]-\varepsilon, \varepsilon[\rightarrow S$ tel que $\gamma(0) = a$ et $\gamma'(0) = v$. Alors, pour tout $t \in]-\varepsilon, \varepsilon[$, $\gamma(t) = (x(t), y(t), z(t))$ avec $z(t) = f(x(t), y(t))$ donc

$$\begin{aligned} z = z'(0) &= x'(0) \frac{\partial f}{\partial x}(x(0), y(0)) + y'(0) \frac{\partial f}{\partial y}(x(0), y(0)) \\ &= x \frac{\partial f}{\partial x}(x_0, y_0) + y \frac{\partial f}{\partial y}(x_0, y_0) \end{aligned}$$

donc

$$0 = x \frac{\partial f}{\partial x}(x_0, y_0) + y \frac{\partial f}{\partial y}(x_0, y_0) - z = (\nabla g(a) | (x, y, z))$$

car $\nabla g(a) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0), -1 \right)$.

■ Si $v = (x, y, z)$ est orthogonal à $\nabla g(a)$, On définit

$$\gamma : \begin{cases}]-1, 1[& \longrightarrow S \\ t & \longrightarrow (x_0 + tx, y_0 + ty, f(x_0 + tx, y_0 + ty)) \end{cases}$$

Alors γ est dérivable en 0, $\gamma(0) = a$ et $\gamma'(0) = \left(x, y, x \frac{\partial f}{\partial x}(x_0, y_0) + y \frac{\partial f}{\partial y}(x_0, y_0) \right) = (x, y, z)$ car $(\nabla g(a) | v) = 0$.

Mais toutes les surfaces n'ont pas une équation explicite $z = f(x, y)$.

Le programme propose un dernier énoncé, à la démonstration hors programme, permettant de traiter le cas général des surfaces décrites par une équation implicite $g(x, y, z) = 0$ et généralisant l'énoncé précédent.



Méthode 1 : Trouver une équation d'hyperplan tangent

On retiendra que dans tous les cas, mieux vaut repasser par une équation **implicite** $g(x) = 0$ pour trouver une équation d'hyperplan tangent.

Le théorème précédent donne alors une équation du-dit hyperplan tangent en un point a n'annulant pas la dg .

Par exemple, le plan tangent à la surface S d'équation $g(x, y, z) = 0$ en $a = (x_0, y_0, z_0)$ est le plan affine $a + \nabla g(a)^\perp$ passant par a et de direction le plan vectoriel $T_a S = \nabla g(a)^\perp$ de vecteur normal $\nabla g(a) \neq (0, 0, 0)$, d'équation

$$\frac{\partial g}{\partial x}(x_0, y_0, z_0)(x - x_0) + \frac{\partial g}{\partial y}(x_0, y_0, z_0)(y - y_0) + \frac{\partial g}{\partial z}(x_0, y_0, z_0)(z - z_0) = 0$$

où $g(a) = g(x_0, y_0, z_0) = 0$.

Exemple

E5 – Équation du plan tangent en un point d'une sphère de \mathbb{R}^3



OPTIMISATION (RECHERCHE D'EXTREMUMS)

Définition 8 : Point critique

Soit $f : \mathcal{U} \rightarrow \mathbb{R}$ différentiable. On dit que a est un **point critique** de f lorsque $df(a) = 0$.

Si E est euclidien, cela équivaut à $\nabla f(a) = 0$.

Propriété 18 : Condition nécessaire d'extremum local

Soit \mathcal{U} **ouvert** (très important) et $f : \mathcal{U} \rightarrow \mathbb{R}$ différentiable en $a \in \mathcal{U}$.

Si f présente un **extremum local** en a , alors a est un **point critique** de f , c'est-à-dire $df(a) = 0$.

Démonstration

Il suffit de revenir aux dérivées partielles.

Propriété 17 : Hyperplan tangent à une surface implicite

Soit g une fonction numérique de classe \mathcal{C}^1 sur l'ouvert \mathcal{U} , $X = \{x \in \mathcal{U}, g(x) = 0_{\mathbb{R}}\}$ et $a \in X$.

Si $dg(a) \neq 0_{\mathcal{L}(E, \mathbb{R})}$, alors

$$T_a X = \text{Ker}(dg(a)) = \nabla g(a)^\perp$$

(L'espace étant supposé euclidien pour utiliser le gradient.)

Autrement dit, X admet un hyperplan affine tangent en a qui est, $a + T_a X = a + \nabla g(a)^\perp$ d'équation

$$(\nabla g(a) | x - a) = 0.$$

Remarque

R13 – Bien sûr le **vecteur** x de cet énoncé ne doit pas être confondu avec le **scalaire** x d'un triplet (x, y, z) du cas particulier \mathbb{R}^3 .

**Propriété 19 : Vecteurs tangents en un extremum local**

Si f est une fonction numérique définie sur l'ouvert \mathcal{U} , X une partie de \mathcal{U} , $f|_X$ admet un extremum local en $a \in X$ et f est différentiable en a , alors

$$\forall v \in T_a X, \quad df(a)(v) = 0.$$

Autrement dit, dans le cas où E est euclidien,

$$T_a X \subset \text{Ker}(df(a)) = (\nabla f(a))^\perp$$

Démonstration

Soit $v \in T_a X$. On a $\gamma :]-\varepsilon, \varepsilon[\rightarrow X$ tel que $\gamma(0) = a$ et $\gamma'(0) = v$.
Alors, $f \circ \gamma$ admet un extremum local en $0 \in]-\varepsilon, \varepsilon[$ qui est ouvert, donc

$$0 = (f \circ \gamma)'(0) = df(\gamma(0))(\gamma'(0)) = df(a)(v).$$

Remarque

R 14 – Autrement dit, les vecteurs tangents à X en a extremum local de $f|_X$ sont dans l'hyperplan vectoriel de vecteur normal $\nabla f(a)$.

Théorème 2 : d'optimisation sous contrainte

Soient $f, g \in \mathcal{C}^1(\mathcal{U}, \mathbb{R})$ où \mathcal{U} est un ouvert de E , et $X = \{x \in \mathcal{U}, g(x) = 0\}$.

Si $f|_X$ admet un extremum local en $a \in X$ et si $dg(a) \neq 0_E$, alors $df(a)$ est colinéaire à $dg(a)$.

Démonstration

- Soit $df(a)$ est nulle et $df(a) = 0 \cdot dg(a)$.
- Soit $df(a)$ n'est pas nulle et on prouve que $\text{Ker}(df(a)) = \text{Ker}(dg(a))$. En effet, un résultat du programme dit que si deux formes linéaires non nulles ont même noyau, elles sont colinéaires.
Comme $dg(a) \neq 0$, et avec la propriété précédente, $\text{Ker}(dg(a)) = T_a X \subset \text{Ker}(df(a))$.
Par égalité des dimensions, $\text{Ker}(df(a)) = \text{Ker}(dg(a))$, ce qui conclut.

Terminons avec une extension hors programme, pour la culture.

Théorème 3 : HP : multiplicateurs de Lagrange

Soit f, g_1, \dots, g_p des fonctions numériques de classe \mathcal{C}^1 sur l'ouvert \mathcal{U} de E et

$$X = \{x \in \mathcal{U}, g_1(x) = \dots = g_p(x) = 0\}.$$

Si $f|_X$ admet un extremum local en $a \in X$ et si les formes linéaires $dg_1(a), \dots, dg_p(a)$ sont linéaires indépendantes, alors il existe $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ tels que

$$df(a) = \lambda_1 dg_1(a) + \dots + \lambda_p dg_p(a).$$

Les réels $\lambda_1, \dots, \lambda_p$ sont appelés **multiplicateurs de Lagrange**.