Funciones de varias variables I

Bibliografía:

- Antonio Avilés.
- FVV1, Luis Oncina (2016–17).

Capítulo 1

Normas y convergencia

Sea E un $\mathbb{R}$-espacio vectorial, una **norma** es una aplicación $\|\cdot\| : E \to \mathbb{R}$ tal que $\forall x, y \in E, \lambda \in \mathbb{R}$:

1. $\|x\| \geq 0 \wedge (\|x\| = 0 \iff x = 0)$.
2. $\|x + y\| \leq \|x\| + \|y\|$.
3. $\|\lambda x\| = |\lambda| \|x\|$.

El par $(E, \|\cdot\|)$ es un **espacio normado**. Llamamos **distancia asociada a la norma** a $d(x, y) := \|x - y\|$. Dos normas son equivalentes si sus distancias lo son.

Ejemplos de normas en $\mathbb{R}^n$ son las dadas por $\|(x_1, \dots, x_n)\|_p := \sqrt[p]{\sum_{i=1}^n |x_i|^p}$ y $\|(x_1, \dots, x_n)\|_\infty := \text{máx}\{|x_i|\}_{i=1}^n$. Además, $V := \mathcal{C}[a, b] := \{f \mid [a, b] \to \mathbb{R} \text{ continua}\}$ con $\|f\|_\infty := \sup\{|f(x)|\}_{x \in [a,b]}$ es un espacio normado.

TEM

$f : (X, d) \to (Y, d')$ es continua en p [...] si y sólo si $\forall \varepsilon > 0, \exists \delta > 0 : \forall x \in X, (d(x, p) < \delta \implies d'(f(x), f(p)) < \varepsilon)$.

Definimos la norma de una aplicación $L : (E, \|\cdot\|) \to (F, \|\cdot\|')$ como $\|L\| := \|L\|_{\|\cdot\|}^{\|\cdot\|'} := \sup\{\|L(x)\|'\}_{x \in E, \|x\| \leq 1}$, y tenemos como **teorema** que L es continua si y sólo si $\|L\| < +\infty$, y entonces es uniformemente continua.

$\implies$] Sea L continua en 0, es decir, $\forall \varepsilon > 0, \exists \delta > 0 : \forall y \in E, (\|y\| < \delta \implies \|L(y)\|' < \varepsilon)$. Fijado ε, sea $\|z\| \leq 1$, entonces $\|\frac{\delta}{2} z\| < \delta$ y $\|L(\frac{\delta}{2} z)\|' < \varepsilon$, luego $\|L(z)\|' < \frac{2\varepsilon}{\delta}$ y $\|L\| < \frac{2\varepsilon}{\delta}$.

$\impliedby$] Veamos primero que $\|L\| < +\infty \implies \forall x \in E, \|L(x)\|' \leq \|L\| \|x\|$. En efecto, para $\|x\| = 1$, $\|L(x)\|' \leq \sup\{\|L(y)\|\}_{\|y\| \leq 1} = \|L\| = \|L\| \|x\|$, y para cualquier otra x basta dividir entre la norma. Ahora bien, dado $\varepsilon > 0$, tomando $\delta := \frac{\varepsilon}{\|L\|+1}$ entonces $\|y - x\| < \delta \implies \|L(y) - L(x)\|' = \|L(y - x)\|' \leq \|L\| \|y - x\| < \|L\| \delta = \frac{\|L\| \varepsilon}{\|L\|+1} < \varepsilon$. Pero como δ no depende de x, L es uniformemente continua.

1.1. Equivalencia de normas

Dos normas $\|\cdot\|$ y $\|\cdot\|'$ son equivalentes si y sólo si $\exists \alpha, \beta > 0 : \forall x \in E, \alpha\|x\| \leq \|x\|' \leq \beta\|x\|$.
Demostración: Sean $L := id_E : (E, \|\cdot\|) \to (E, \|\cdot\|')$ y $L' := L^{-1}$, entonces $\mathcal{T}_{\|\cdot\|} = \mathcal{T}_{\|\cdot\|'}$ si y sólo si L y L' son continuas, pues L es continua $\iff \forall A \in \mathcal{T}_{\|\cdot\|'}, A \in \mathcal{T}_{\|\cdot\|} \iff \mathcal{T}_{\|\cdot\|'} \subseteq \mathcal{T}_{\|\cdot\|}$, y el otro contenido es análogo. Entonces:

$\Longrightarrow$] Si L es continua $\|L\| < +\infty$, luego $\|x\|' = \|L(x)\|' \leq \|L\|\|x\| \overset{\beta := \|L\|}{=} \beta\|x\|$. La otra cota se hace de forma análoga.

$\Longleftarrow$] Si existe β tal que $\forall x \in E, \|x\|' \leq \beta\|x\|$, en particular se cumple para $\|x\| \leq 1$, y entonces $\|L(x)\| = \|x\|' \leq \beta\|x\|$, luego $\|L\| = \sup\{\|x\|'\}_{\|x\| \leq 1} \leq \beta < +\infty$ y L es continua.

TEM

Las métricas d_E, d_T y d_∞ [...] son equivalentes, [...]. **Demostración:** Se deduce de que $\frac{1}{n}d_T(x,y) \leq d_\infty(x,y) \leq d_T(x,y)$ y $\frac{1}{\sqrt{n}}d_E(x,y) \leq d_\infty(x,y) \leq d_E(x,y)$. [...]
Todo cerrado C de un compacto $(X, \mathcal{T})$ es compacto. [...] En [...] $(\mathbb{R}, \mathcal{T}_u)$ [...] todo subespacio cerrado y acotado es compacto. [...] Todo subespacio compacto K de un espacio métrico (X, d) es [cerrado y] acotado. [...]Si $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ es continua y $(X, \mathcal{T})$ es compacto entonces $f(X)$ es compacto. [...] $(X, \mathcal{T})$ es **compacto** [...] si toda sucesión admite una subsucesión convergente.

Toda norma $\|\cdot\| : (E, \|\cdot\|) \to \mathbb{R}$ es uniformemente continua. **Demostración:** Fijado ε y tomando $\delta := \varepsilon$, si $\|x - y\| < \delta$ entonces usando que $|\|x\| - \|y\|| \leq \|x - y\|$, lo que se deduce de $\|x\| \leq \|x - y\| + \|y\|$ y $\|y\| \leq \|y - x\| + \|x\|$, obtenemos que $|\|x\| - \|y\|| < \varepsilon$.
Como **teorema**, en $\mathbb{R}^n$, todas las normas son equivalentes.

$\|x\| \leq C\|x\|_1$] $\|x\| = \|\sum x_i \vec{e}_i\| \leq \sum |x_i|\|\vec{e}_i\| \leq \text{máx}\{\|\vec{e}_i\|\} \sum |x_i| = \text{máx}\{\|\vec{e}_i\|\}\|x\|_1$.

$\|x\|_1 \leq D\|x\|$] Tomamos $(\mathbb{R}^n, \|\cdot\|_1) \overset{id}{\to} (\mathbb{R}^n, \|\cdot\|) \overset{\|\cdot\|}{\to} \mathbb{R}$, que es continua por ser composición de dos funciones continuas (la identidad es continua por la otra cota y la demostración del teorema anterior), entonces $S := \{x \in \mathbb{R}^n \mid \|x\|_1 = 1\}$ es cerrado dentro del compacto $\overline{B}(0,1)$, luego es compacto y como la función dada es continua, $\|\cdot\|(S)$ es compacto y alcanza su máximo y su mínimo. Sea ahora $\mu := \text{mín}\{\|x\|\}_{x \in S} > 0$ (pues $0 \notin S$), si $\|x\|_1 = 1$ para un cierto $x \in \mathbb{R}^n$, entonces $\|x\| \geq \mu$, luego $x \neq 0$ y $\left\|\frac{x}{\|x\|_1}\right\| = 1$, con lo que $\left\|\frac{x}{\|x\|_1}\right\| \geq \mu$ y $\|x\| \geq \mu\|x\|_1$, y entonces $\|x\|_1 \leq \frac{1}{\mu}\|x\|$.

Tenemos pues que toda $T : (\mathbb{R}^m, \|\cdot\|) \to (\mathbb{R}^n, \|\cdot\|')$ lineal es continua

1.2. Convergencia

Sea $(X, \mathcal{T})$ un espacio topológico e (Y, d) un espacio métrico, una sucesión de funciones $(f_n : (X, \mathcal{T}) \to (Y, d))_n$ **converge puntualmente** a $f : (X, \mathcal{T}) \to (Y, d)$ si para todo $x \in X$, $f_n(x) \to f(x)$, y converge **uniformemente** a f si $d_\infty(f_n, f) \to 0$. Sea $x_0 \in (X, \mathcal{T})$ y $(f_n : X \to \mathbb{R})_n$ una sucesión de funciones continuas en x_0 que converge uniformemente a f, entonces f es continua en x_0. **Demostración:** Fijado un $\varepsilon > 0$, existe un $n_0 \in \mathbb{N}$ tal que para $n \geq n_0$ y

$x \in X$ se tiene $|f_n(x) - f(x)| < \frac{\varepsilon}{3}$. Como f_n es continua, existe $\mathcal{V} \in \mathcal{E}(x_0)$ tal que si $x \in \mathcal{V}$ entonces $|f_n(x) - f_n(x_0)| < \frac{\varepsilon}{3}$, con lo que $|f(x) - f(x_0)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$.

FUVR1

1. Si $\lim_n x_n = \pm\infty$ entonces $\lim_n \left(1 + \frac{1}{x_n}\right)^{x_n} = e$ y $\lim_n \left(1 - \frac{1}{x_n}\right)^{x_n} = e^{-1}$.

2. Si existe $\lim_n \frac{z_{n+1}}{z_n} = w \in \mathbb{R}$ con $|w| < 1$, entonces $\lim_n z_n = 0$. [...]

Si $b > 0$, $c > 1$ y $d > 0$, entonces

$$\log n \ll n^b \ll c^n \ll n^{dn}$$

Si además $d \geq 1$, entonces $c^n \ll n! \ll n^{dn}$. [...]

Si $\lim_n x_n = 0$ con $0 < |x_n| < 1$, entonces:

1. $\log(1 + x_n) \sim x_n$.

2. $e^{x_n} - 1 \sim x_n$.

Si $\lim_n x_n = 1$ con $x_n \neq 1$ y $\lim_n y_n = \pm\infty$, entonces

$$\lim_n x_n^{y_n} = e^{\lim_n y_n (x_n - 1)}$$

Si $\lim_n x_n = 0$ y $x_n \neq 0$, entonces $\sin x_n \sim x_n$.

Criterios de Stolz: Si $(a_n)_n$ y $(b_n)_n$ son sucesiones de reales tales que $(b_n)_n$ es estrictamente creciente o decreciente y bien $\lim_n a_n = \lim_n b_n = 0$, bien $\lim_n b_n = \infty$, si existe $\lim_n \frac{a_{n+1} - a_n}{b_{n+1} - b_n} = L \in \overline{\mathbb{R}}$, entonces $\lim_n \frac{a_n}{b_n} = L$.

Como consecuencia:

1. Si $(a_n)_n$ converge, entonces

$$\lim_n \frac{a_1 + \cdots + a_n}{n} = \lim_n a_n$$

2. Si $(a_n)_n$ converge y $a_n > 0$, entonces

$$\lim_n \sqrt[n]{a_1 \cdots a_n} = \lim_n a_n$$

3. Si $a_n > 0$ y existe $\lim_n \frac{a_n}{a_{n-1}}$, entonces

$$\lim_n \sqrt[n]{a_n} = \lim_n \frac{a_n}{a_{n-1}}$$

FUVR1

La **condición de Cauchy** nos dice que $\sum_n a_n$ es convergente si y sólo si $\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : \forall p, q \in \mathbb{N}, (n_0 \leq p \leq q \implies |a_{p+1} + \cdots + a_q| < \varepsilon)$.

[...] Si S_n converge, entonces $\lim_n a_n = 0$. [...] La convergencia de una serie no se altera modificando un número finito de términos de esta. [...]

Dada una serie $\sum_{n=1}^{\infty} a_n$ de términos $a_n \geq 0$, esta es convergente si y sólo si la sucesión de sumas parciales es acotada [...].

Criterios de comparación:

1. Dadas $\sum_n a_n$ y $\sum_n b_n$ con $a_n, b_n \geq 0$, si existe $M > 0$ tal que $a_n \leq M b_n \forall n$, entonces la convergencia de $\sum_{n=1}^{\infty} b_n$ implica la de $\sum_{n=1}^{\infty} a_n$ [...].

2. Dadas $\sum_n a_n$ y $\sum_n b_n$ con $a_n, b_n > 0$ y existe $l := \lim_n \frac{a_n}{b_n} \in \mathbb{R} \cup \{+\infty\}$:

   a) Si $0 < l < \infty$, ambas series tienen el mismo carácter. [...]

   b) Si $l = 0$ entonces la convergencia de $\sum_n b_n$ implica la de $\sum_n a_n$. [...]

   c) Si $l = +\infty$ entonces la convergencia de $\sum_n a_n$ implica la de $\sum_n b_n$. [...]

Criterio de la raíz: Dada $\sum_n a_n$ con $a_n > 0$ y $a := \lim_n \sqrt[n]{a_n} \in \mathbb{R}$:

- Si $a < 1$, la serie converge. [...]
- Si $a > 1$, la serie diverge. [...]
- Si $a = 1$ no se puede afirmar nada.

Criterio del cociente: Sea $\sum_n a_n$ con $a_n > 0$ y $a := \lim_n \frac{a_{n+1}}{a_n} \in \mathbb{R}$. [...]

- Si $a < 1$, la serie converge.
- Si $a > 1$, la serie diverge.

Criterio de condensación: Dada una sucesión $(a_n)_n$ monótona decreciente con $a_n > 0$. Entonces

$$\sum_{n=1}^{\infty} a_n \in \mathbb{R} \iff \sum_{n=1}^{\infty} 2^n a_{2^n} \in \mathbb{R}$$

Una serie $\sum_n a_n$ con $a_n \in \mathbb{R}$ es **absolutamente convergente** si $\sum_n |a_n|$ es convergente. Toda serie absolutamente convergente es convergente. [...]

La **serie geométrica** $\sum_{n=0}^{\infty} r^n$ es convergente si $|r| < 1$ con suma $\frac{1}{1-r}$ y divergente si $|r| \geq 1$. La **serie armónica** $\sum_{n=1}^{\infty} \frac{1}{n^k}$ es convergente si $k > 1$ y divergente si $k \leq 1$.

1.3. Completitud

Una sucesión (x_n) en un espacio métrico (E, d) es **de Cauchy** si $\forall \varepsilon > 0, \exists n_\varepsilon \in \mathbb{N} : \forall n, m \geq n_\varepsilon, d(x_n, x_m) < \varepsilon$. Un espacio métrico es **completo** si toda sucesión de Cauchy es convergente. Un **espacio de Banach** es un espacio normado completo. Dadas dos normas $\|\cdot\|$ y $\|\cdot\|'$ equivalentes sobre E, $(E, \|\cdot\|)$ es completo si y sólo si lo es $(E, \|\cdot\|')$.

Como **teorema**, $\mathbb{R}^n$ es un espacio de Banach con cualquier norma. **Demostración:** Basta probar que $(\mathbb{R}^n, \|\cdot\|_\infty)$ es completo. Si (x_m) es de Cauchy en $(\mathbb{R}^n, \|\cdot\|_\infty)$, como $|x_{mi} - x_{ki}| \leq \|x_m - x_k\|_\infty$ para todo $i \in \{1, \ldots, n\}$, entonces $(x_{mi})_m$ es de Cauchy en $\mathbb{R}$ y por tanto convergente a un x_{0i}, con lo que $(x_m)_m$ converge a $(x_{01}, \ldots, x_{0n})$, y se tiene que $(\mathbb{R}^n, \|\cdot\|)$ es completo.

Como **teorema**, $(\mathcal{C}[a,b], \|\cdot\|_\infty)$ es un espacio de Banach. **Demostración:** Sea $(f_n)_n$ una sucesión de Cauchy en $(\mathcal{C}[a,b], \|\cdot\|_\infty)$, fijado un $\varepsilon > 0$ existe un n_0 tal que para $n, m \geq n_0$ y $x \in [a,b]$, $|f_n(x) - f_m(x)| < \frac{\varepsilon}{2}$. Por tanto para cada $x \in [a,b]$, $(f_n(x))_n$ es de Cauchy en $\mathbb{R}$ y converge pues a un valor $f(x)$. Ahora bien, dado un $x \in [a,b]$, por la convergencia puntual que acabamos de probar existe $n_x \in \mathbb{N}$ tal que para $n \geq n_x$ se tiene $|f_n(x) - f(x)| < \frac{\varepsilon}{2}$. Si $n > n_0$, $|f_n(x) - f(x)| \leq |f_n(x) - f_{\text{máx}\{n_0,n_x\}}(x)| + |f_{\text{máx}\{n_0,n_x\}}(x) - f(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, y como n_0 no depende de x (sólo de ε), queda probada la convergencia uniforme.

Sea $(x_n)_n$ una sucesión en el espacio de Banach $(E, \|\cdot\|)$ con $\sum_n \|x_n\| < +\infty$, entonces $\sum_n x_n$ converge.

Capítulo 2

Derivadas y diferenciales

2.1. Derivadas

Sean E y F normados, $\Omega \subseteq E$ abierto y $f : \Omega \to F$, dados $a \in \Omega$ y $u \in E$, f es **derivable** en a según u si existe la **derivada direccional** de f en x según u, dada por

$$d_u f(x_0) := \lim_{t \to 0} \frac{f(a+tu) - f(a)}{t}$$

Si $u = \vec{e}_i$ es el vector i-ésimo de la base canónica hablamos de la **derivada parcial** i-ésima, que denotamos

$$\frac{\partial f}{\partial x_i}(a) := d_i f(a) := d_{\vec{e}_i} f(a)$$

Si $\vec{u} = \lambda \vec{v}$ con $\lambda \in \mathbb{R}$, $d_{\vec{u}} f(t)$ existe si y sólo si existe $d_{\vec{v}} f(t)$.

$f : (a, b) \to \mathbb{R}^n$ es derivable en x_0 si cada una de sus coordenadas lo es, y entonces $f'(x_0) = (f_1'(x_0), \ldots, f_n'(x_0))$.

2.2. Diferenciales

Sea $\Omega \subseteq E$ abierto, $f : \Omega \to F$ y $a \in \Omega$, f es **diferenciable** en a si existe una aplicación lineal $L : E \to F$ tal que

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - L(h)}{\|h\|} = 0$$

Esta aplicación es la **diferencial** de f en a, denotada por $df(a)$, y si existe es única.

Escribimos $L \equiv M$ si M es la matriz asociada a la aplicación lineal L.

$f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n$ es diferenciable en $a \in \Omega$ con diferencial L si y sólo si cada $f_i : \mathbb{R}^m \to \mathbb{R}$, $i \in \{1, \ldots, n\}$, es diferenciable con diferencial L_i.

Si $f : \Omega \subseteq E \to F$ es diferenciable en $a \in \Omega$, también es continua en a.

$f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n$ es diferenciable en $a \in \Omega$ (Ω abierto) si para todo $P \in \Omega$ existen todas las derivadas parciales $\frac{\partial f_i}{\partial x_j}(P)$ y son continuas en a.

Demostración: Podemos suponer $n=1$, pues de lo contrario basta probar que cada f_i es diferenciable en a. Se trata pues de probar que $\lim_{h\to 0}\frac{f(a+h)-f(a)-\sum_{i=1}^m \frac{\partial f}{\partial x_i}(a)h_i}{\|h\|}=0$, lo que ocurre si y sólo si

$$\begin{aligned}
0 &= \lim_{h\to 0}\frac{|f(a+h)-f(a)-\sum_{i=1}^m \frac{\partial f}{\partial x_i}(a)h_i|}{\|h\|} = \lim_{h\to 0}\frac{f(a+\sum_{i=1}^m h_i\vec{e}_i)-f(a)-\sum_{i=1}^m \frac{\partial f}{\partial x_i}(a)h_i}{\|h\|} \\
&= \lim_{h\to 0}\frac{\left|\sum_{i=1}^m \left(f(a+h_1\vec{e}_1+\cdots+h_i\vec{e}_i)-f(a+h_1\vec{e}_1+\cdots+h_{i-1}\vec{e}_{i-1})-\frac{\partial f}{\partial x_i}(a)h_i\right)\right|}{\|h\|}
\end{aligned}$$

El último sumatorio con sus dos primeros elementos forma una **suma telescópica**: todos los elementos se anulan salvo el primero y el último. Sabemos que cada $a+h_1\vec{e}_1+\cdots+h_i\vec{e}_i$ está en el dominio de f porque Ω es abierto y h se supone lo suficientemente pequeño. Ahora llamamos $\varphi_i(t)\coloneqq f(a+h_1\vec{e}_1+\cdots+h_{i-1}\vec{e}_{i-1}+t\vec{e}_i)$, con lo que $\varphi_i'(t)=\frac{\partial f}{\partial x_i}(a+h_1\vec{e}_1+\cdots+h_{i-1}\vec{e}_{i-1}+t\vec{e}_i)$, y $\Delta_i\coloneqq\varphi_i(h_i)-\varphi_i(0)=f(a+h_1\vec{e}_1+\cdots+h_i\vec{e}_i)-f(a+h_1\vec{e}_1+\cdots+h_{i-1}\vec{e}_{i-1})=\varphi_i'(\xi_i)h_i$ para algún ξ_i entre 0 y h_i, que tiende a 0. Sustituyendo nos queda que lo anterior es igual a

$$\lim_{h\to 0}\frac{\left|\sum_{i=1}^m \varphi_i'(\xi_i)h_i-\frac{\partial f}{\partial x_i}(a)h_i\right|}{\|h\|}$$

Entonces,

$$\begin{aligned}
0 &\le \frac{\left|\sum_{i=1}^m \varphi_i'(\xi_i)h_i-\frac{\partial f}{\partial x_i}(a)h_i\right|}{\|h\|_\infty} \\
&\le \left|\sum_{i=1}^m \frac{\partial f}{\partial x_i}(a+h_1\vec{e}_1+\cdots+h_{i-1}\vec{e}_{i-1}+\xi_i\vec{e}_i)-\frac{\partial f}{\partial x_i}(a)\right|\frac{|h_i|}{\|h\|_\infty}\to 0
\end{aligned}$$

Que esta última expresión tienda a 0 se debe a que $0\le\frac{|h_i|}{\|h\|_\infty}\le 1$ y a que las derivadas parciales de f sean continuas y por tanto $\lim_{h\to 0}\frac{\partial f}{\partial x_i}(a+\dots)=\frac{\partial f}{\partial x_i}(\lim_{h\to 0}(a+\dots))$. Entonces, por la regla del sandwich, el límite inicial tiende a 0. Hemos utilizado la norma $\|\cdot\|_\infty$, pero como dada una norma $\|\cdot\|$ $\exists\alpha,\beta>0:\forall h,\alpha\le\frac{\|h\|_\infty}{\|h\|}\le\beta$, la convergencia a 0 no depende de la norma que tomemos.

2.3. Regla de la cadena

La **regla de la cadena** afirma que si $\mathcal{U}\subseteq\mathbb{R}^m$ y $\mathcal{V}\subseteq\mathbb{R}^n$ son abiertos, $a\in\mathcal{U}$ y $\mathcal{U}\xrightarrow{f}\mathcal{V}\xrightarrow{g}\mathbb{R}^k$ son diferenciables en a y en $f(a)$, respectivamente, entonces $g\circ f$ es diferenciable en a y $d(g\circ f)(a)=dg(f(a))\circ df(a)$.

Demostración: Sean $L\coloneqq df(a):\mathbb{R}^m\to\mathbb{R}^n$ y $S\coloneqq dg(f(a)):\mathbb{R}^n\to\mathbb{R}^k$, tenemos que

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-L(h)}{\|h\|}=\lim_{\eta\to 0}\frac{g(f(a)+\eta)-g(f(a))-S(\eta)}{\|\eta\|}=0$$

y queremos ver que

$$\lim_{h\to 0}\frac{g(f(a+h))-g(f(a))-S(L(h))}{\|h\|}=0$$

Si llamamos $\eta \coloneqq f(a+h)-f(a)$, que tiende a 0 por la continuidad de f en a, entonces

$$\begin{aligned}\lim_{h\to 0} \frac{g(f(a+h))-g(f(a))-S(L(h))}{\|h\|} &= \lim_{h\to 0} \frac{g(f(a)+\eta)-g(f(a))-S(\eta)}{\|h\|} + \frac{S(\eta)-S(L(h))}{\|h\|} \\ &= \lim_{h\to 0} \frac{g(f(a)+\eta)-g(f(a))-S(\eta)}{\|\eta\|}\frac{\|\eta\|}{\|h\|} - S\left(\frac{f(a+h)-f(a)-L(h)}{\|h\|}\right)\end{aligned}$$

Como $S\left(\frac{f(a+h)-f(a)-L(h)}{\|h\|}\right) \to 0$ usando la linealidad de S y su continuidad (que se deduce de su linealidad), y como $\frac{g(f(a)+\eta)-g(f(a))-S(\eta)}{\|\eta\|} \to 0$, el límite tenderá a 0 si y sólo si $\frac{\|\eta\|}{\|h\|}$ es acotado, pero

$$\begin{aligned} 0 &\leq \frac{\|\eta\|}{\|h\|} = \frac{\|f(a+h)-f(a)-L(h)+L(h)\|}{\|h\|} \\ &\leq \frac{\|f(a+h)-f(a)-L(h)\|}{\|h\|} + \frac{\|L(h)\|}{\|h\|} \to 0 + \frac{\|L(h)\|}{\|h\|} \leq \frac{\|L\|\|h\|}{\|h\|} < +\infty \end{aligned}$$

por la continuidad de L.

2.4. Incremento finito

El **teorema del incremento finito** afirma que, sean $f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n$, $a, b \in \Omega$ con el segmento $[a,b] \subseteq \Omega$ y $M > 0$, si $\|df(x)\| \leq M$ para todo $x \in [a,b]$ se tiene $\|f(b)-f(a)\| \leq M\|b-a\|$.

Demostración: Fijado $\varepsilon > 0$, sabemos que para $x \in [a,b]$, $\lim_{h\to 0} \frac{f(x+h)-f(x)-df(x)(h)}{\|h\|} = 0$ y por tanto existe $\delta_x > 0$ tal que para $\|h\| < \delta_x$ se tiene

$$\|f(x+h)-f(x)-df(x)(h)\| < \varepsilon\|h\|$$

con lo que

$$\|f(x+h)-f(x)\| - \|df(x)(h)\| \leq \|f(x+h)-f(x)-df(x)(h)\| < \varepsilon\|h\|$$

y por tanto

$$\|f(x+h)-f(x)\| < \varepsilon\|h\| + \|df(x)(h)\| \leq \varepsilon\|h\| + \|df(x)\|\|h\| \leq (\varepsilon+M)\|h\|$$

Esta desigualdad depende de δ_x y por tanto de x. Sea entonces $\{B(x, \frac{\delta_x}{2})\}_{x\in[a,b]}$ un recubrimiento por abiertos de $[a,b]$ y $\{B_i\}_{i=1}^k \coloneqq \{B(x_i, \frac{\delta_{x_i}}{2})\}_{i=1}^k$ un subrecubrimiento finito del que suponemos que no podemos quitar ninguna bola. Ahora llamamos $x_0 \coloneqq a$ y $x_{k+1} \coloneqq b$ y suponemos $a = x_0 < x_1 < \cdots < x_k < x_{k+1} = b$. Por la desigualdad anterior, para $x, y \in [a,b]$ con $\|y-x\| < \delta_x$ o $\|x-y\| < \delta_y$, $\|f(y)-f(x)\| \leq (M+\varepsilon)\|x-y\|$. El segmento $[x_i, x_{i+1}]$ queda cubierto por B_i y B_{i+1}, pues si hiciera falta además B_j con $j \neq i, i+1$ para cubrirlo sería $x_j < x_i$ y entonces $B_i \subseteq B_j$ o $x_j > x_{i+1}$ y entonces $B_{i+1} \subseteq B_j$, pero entonces podríamos quitar una bola del recubrimiento#. Por tanto $\|x_{i+1}-x_i\| < \frac{\delta_{x_i}}{2} + \frac{\delta_{x_{i+1}}}{2} \leq \max\{\delta_{x_i}, \delta_{x_{i+1}}\}$. Finalmente tenemos que $\|f(b)-f(a)\| = \|f(x_{k+1})-f(x_k)+\cdots+f(x_1)-f(x_0)\| \leq \sum_{i=0}^k \|f(x_{i+1})-f(x_i)\| \leq \sum_{i=0}^k \|x_{i+1}-x_i\|(M+\varepsilon)$ y, como todos los $x_{i+1}-x_i$ tienen la forma $\lambda(b-a)$ con $\lambda > 0$, entonces $\sum_{i=0}^k \|x_{i+1}-x_i\|(M+\varepsilon) = \|b-a\|(M+\varepsilon)$. Como esto se da para todo $\varepsilon > 0$, el resultado queda probado.

Capítulo 3

Dobles diferenciales

$f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n$ es **dos veces diferenciable** o **de clase** $\mathcal{C}^2$ en $a \in \Omega$ si f es diferenciable en $\mathcal{U} \in \mathcal{E}(a)$ y $df : \mathcal{U} \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \equiv M_{n\times m}(\mathbb{R}) \equiv \mathbb{R}^{nm}$ (la aplicación que a cada elemento de $\mathcal{U}$ le asigna un vector en $\mathbb{R}^{nm}$ que contiene, en algún orden, los elementos de la matriz asociada a la diferencial del elemento) es diferenciable en a. Por inducción se define el ser n veces diferenciable o de clase $\mathcal{C}^n$, y el ser infinitamente diferenciable o de clase $\mathcal{C}^\infty$.

Denotamos la derivada parcial k-ésima de la derivada parcial j-ésima de la i-ésima coordenada de f, o la i-ésima coordenada de la doble derivada parcial respecto a x_j y x_k, como $\frac{\partial^2 f_i}{\partial x_j \partial x_k}$, y si $j = k$, también escribimos $\frac{\partial^2 f_i}{\partial x_j^2}$. Si f tiene derivadas parciales segundas $\frac{\partial f_i}{\partial x_k \partial x_j}$ todas continuas en a entonces f es dos veces diferenciable en a.

3.1. Matriz hessiana

Del mismo modo que podemos pensar en la diferencial de una función diferenciable como $df(a) : \mathbb{R}^m \to \mathbb{R}^n$ dada por $\vec{u} \mapsto d_{\vec{u}} f(a)$, llamamos **diferencial segunda** de f en a a la aplicación $d^2 f(a) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n$ con $(\vec{u}, \vec{v}) \mapsto d_{\vec{v}} d_{\vec{u}} f(a)$, y vemos que esta es una aplicación bilineal.

La matriz de $d^2 f(a) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, dada por

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{ij}$$

se denomina **matriz hessiana**. Así, si M es la matriz hessiana de f, entonces

$$d^2 f(a)(\vec{u}, \vec{v}) = \begin{pmatrix} - & \vec{u} & - \end{pmatrix} M \begin{pmatrix} | \\ \vec{v} \\ | \end{pmatrix}$$

Como **teorema**, sea $f : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ y $a = (x_0, y_0) \in \Omega$, si $\frac{\partial^2 f}{\partial x \partial y}$ y $\frac{\partial^2 f}{\partial y \partial x}$ están definidas en Ω y son continuas en a, entonces su valor en a coincide. Esto significa que la matriz Hessiana es simétrica. **Demostración:** Como Ω es abierto, existe ε tal que $B_\infty(a, \varepsilon) = (x_0 - \varepsilon, x_0 + \varepsilon) \times (y_0 - \varepsilon, y_0 + \varepsilon) \subseteq \Omega$. Fijamos $t \in (x_0 - \varepsilon, y_0 + \varepsilon)$ y $s \in (y_0 - \varepsilon, y_0 + \varepsilon)$, y consideramos

$\Delta_{t,s} \coloneqq f(t,s) - f(t,y_0) - f(x_0,s) + f(x_0,y_0)$. Si ahora llamamos $F_{\overline{s}}(\overline{t}) \coloneqq f(\overline{t},\overline{s}) - f(\overline{t},y_0)$, vemos que $F_{\overline{s}}(\overline{t})$ es derivable con $F'_{\overline{s}}(\overline{t}) = \frac{\partial f}{\partial x}(\overline{t},\overline{s}) - \frac{\partial f}{\partial x}(\overline{t},y_0)$ y que entonces $\Delta_{t,s} = F_s(t) - F_s(x_0) = F'_{\overline{s}}(\xi_{t,s})(t-x_0) = \left(\frac{\partial f}{\partial x}(\xi_{t,s},s) - \frac{\partial f}{\partial x}(\xi_{t,s},y_0)\right)(t-x_0) \overset{\Phi(\overline{s}) \coloneqq \frac{\partial f}{\partial x}(\xi_{t,s},\overline{s})}{=} (\Phi(s)-\Phi(y_0))(t-x_0) \overset{\Phi \text{ derivable por hipótesis}}{=} \Phi'(\eta_{t,s})(s-y_0)(t-x_0) = \frac{\partial^2 f}{\partial x \partial y}(\xi_{t,s},\eta_{t,s})(s-y_0)(t-x_0)$. Permutando los papeles de las dos coordenadas (definiendo $\sigma_{\overline{t}}(\overline{s}) \coloneqq f(\overline{t},\overline{s}) - f(x,\overline{s})$) obtenemos que $\Delta_{t,s} = \sigma_{\overline{t}}(\overline{s}) - \sigma_{\overline{t}}(y_0) = \frac{\partial^2 f}{\partial y \partial x}(\hat{\xi}_{t,s},\hat{\eta}_{t,s})(s-y_0)(t-x_0)$. Haciendo ahora tender (t,s) a (x_0,y_0), por la regla del sandwich $(\xi_{t,s},\eta_{ts})$ y $(\hat{\xi}_{t,s},\hat{\eta}_{t,s})$ también tienden a (x_0,y_0), y aplicando la continuidad de las derivadas parciales dobles en a, nos queda finalmente que $\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0) = \frac{\partial^2 f}{\partial y \partial x}(x_0,y_0)$.

3.2. Desarrollos de Taylor

Despejando de la definición de diferencial, nos queda que $f(a+h) = f(a) + df(a)(h) + o(\|h\|)$, lo que podemos interpretar como una aproximación de $f(x)$ cerca de a por un polinomio de grado 1. Como **teorema**, si $f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}$ es dos veces diferenciable en $a \in \Omega$ entonces $f(a+h) = f(a) + df(a)(h) + \frac{1}{2}d^2f(a)(h,h) + o(\|h\|^2)$.

Demostración: Sea $R(h) \coloneqq f(a+h) - f(a) - df(a)(h) - \frac{1}{2}d^2f(a)(h,h)$, y hemos de ver que $\lim_{h\to 0}\frac{R(h)}{\|h\|^2} = 0$. Como todas las normas en $\mathbb{R}^m$ son equivalentes, elegimos $\|\cdot\|_\infty$. Usamos el teorema del incremento finito, que afirma que si R es diferenciable y $\|dR(\xi)\| \leq M \forall \xi \in [0,h]$ entonces $\|R(h)-R(0)\| \leq M \cdot \|h-0\|$. R es diferenciable al ser la suma de $f(c+h)$ y un polinomio de grado máximo 2. Para estimar $\|dR\|$ vemos que $R(a) = f(a+h) - f(a) - \sum_i \frac{\partial f}{\partial x_i}(a)h_i - \frac{1}{2}\sum_{i,j}\frac{\partial f}{\partial x_i \partial x_j}(a)h_ih_j$, y usando la δ de Kronecker,

$$\frac{\partial}{\partial x_k}\frac{1}{2}\sum_{i,j}\frac{\partial f}{\partial x_i \partial x_j}(a)h_ih_j = \frac{1}{2}\sum_{i,j}\left(\frac{\partial f}{\partial x_i x_j}(a)\delta_{ik} + \frac{\partial f}{\partial x_i \partial x_j}(a)\delta_{jk}\right) = \frac{1}{2}\cdot 2d\left(\frac{\partial f}{\partial x_k}\right)(a)(h)$$

Por tanto

$$\frac{\partial R}{\partial x_k} = \frac{\partial f}{\partial x_k}(a+h) - 0 - \frac{\partial f}{\partial x_k}(a) - d\left(\frac{\partial f}{\partial x_k}\right)(a)(h) =: \psi_k(h)\|h\|$$

donde $\lim_{h\to 0}\psi(h) = 0$. Como $\frac{\partial R}{\partial x_k}$ es continua, definiendo el compacto $[0,h]$ como $\{th\}_{t\in[0,1]}$ existe un punto $t_{k,h}h \in [0,h]$ tal que

$$\frac{\partial R}{\partial x_k}(t_{k,h}h) = \max\left\{\frac{\partial R}{\partial x_k}(\xi)\right\}_{\xi\in[0,h]}$$

Por esto, y como $dR(\xi) \equiv \left(\frac{\partial R}{\partial x_1}(\xi), \dots, \frac{\partial R}{\partial x_m}(\xi)\right)$, existe C tal que

$$\|dR(\xi)\| \leq C \cdot \|dR(a)\|_\infty = C \cdot \max\left\{\left|\frac{\partial R}{\partial x_k}(a)\right|\right\}_{k=1}^n$$

para $\xi \in [0,h]$, y por el teorema del incremento finito, si p es tal que

$$\left|\frac{\partial R}{\partial x_p}(a)\right| = \max\left\{\left|\frac{\partial R}{\partial x_k}(a)\right|\right\}_{k=1}^n$$

tenemos

$$\|R(h)\| = \|R(h) - R(0)\| \le C \left| \frac{\partial R}{\partial x_p}(t_{p,h}) \right| \|h\| = C\psi_p(t_{p,h}h)\|h\|\|h\| = C\psi_p(t_{p,h}h)\|h\|^2$$

y entonces $\frac{|R(h)|}{\|h\|^2} \le C\psi_p(t_{p,h}h) \to 0$, lo que prueba el teorema.

3.3. Extremos relativos

Si V es un K-espacio vectorial con $k := \dim_K(V) < +\infty$ y $\sigma : V \times V \to \mathbb{R}$ una aplicación bilineal, existe $A = (a_{ij}) \in \mathcal{M}_k(K)$ asociado a σ y podemos definir

$$\Delta_1 = |a_{11}|, \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \dots, \Delta_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix}$$

Entonces un **teorema** de álgebra nos dice que σ es:

1. Semidefinida positiva si y sólo si $\Delta_i \ge 0 \forall i$.
2. Semidefinida negativa si y sólo si $\Delta_i(-1)^i \ge 0 \forall i$.
3. Definida positiva si y sólo si $\Delta_i > 0 \forall i$.
4. Definida negativa si y sólo si $\Delta_i(-1)^i > 0 \forall i$.

Como **teorema**, sea $\Omega \subseteq \mathbb{R}^m$ abierto, $f : \Omega \to \mathbb{R}$ y $a \in \Omega$,

1. Si f alcanza en a un extremo relativo entonces $df(a) = 0$.
   Podemos suponer que alcanza un máximo. Entonces $\exists \mathcal{U} \in \mathcal{E}(a) : f(x) \le f(a) \forall x \in \mathcal{U}$, luego si para $i \in \{1, \dots, m\}$ definimos $\varphi_i(t) := f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_m)$, fijado i, $\exists \varepsilon > 0 : \forall t \in (a - \varepsilon, a + \varepsilon), \varphi_i(t) \le \varphi_i(a_i)$ y $\frac{\partial f}{\partial x_i}(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_m) = \varphi'(t)$, luego $\frac{\partial f}{\partial x_i}(a) = \varphi'(a_i) = 0$ y

$$df(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) & \cdots & \frac{\partial f}{\partial x_m}(a) \end{pmatrix} = 0$$

2. Si f es de clase $\mathcal{C}^2$ y $df(a) = 0$, entonces

$$\begin{aligned} d^2 f(a) \text{ definida positiva} &\implies f \text{ tiene un mínimo estricto en } a \implies \\ \implies f \text{ tiene un mínimo en } a &\implies d^2 f(a) \text{ semidefinida positiva} \end{aligned}$$

   $1 \implies 2]$ Consideremos el desarrollo de Taylor de f de orden 2 en a, que como $df(a) = 0$, queda como

$$\begin{aligned} f(x) &= f(a) + \frac{1}{2} d^2 f(a)(x-a, x-a) + o(\|x-a\|^2) \\ &= f(a) + \frac{1}{2}\|x-a\|^2 d^2 f(a)\left(\frac{x-a}{\|x-a\|}, \frac{x-a}{\|x-a\|}\right) + o(\|x-a\|^2) \\ &= f(a) + \frac{1}{2}\|x-a\|^2 \left(d^2 f(a)\left(\frac{x-a}{\|x-a\|}, \frac{x-a}{\|x-a\|}\right) + \frac{2o(\|x-a\|^2)}{\|x-a\|^2}\right) \end{aligned}$$

suponiendo $x \neq a$. Pero $\frac{x-a}{\|x-a\|} \in \{y \in \mathbb{R}^m \mid \|y\| = 1\} =: K$, que es compacto por ser cerrado y acotado, y $\Phi : \mathbb{R}^m \to \mathbb{R}$ dada por $\Phi(u) := d^2f(a)(u,u) = \sum_{i,j} \frac{\partial f}{\partial x_i \partial x_j}(a)u_i u_j$ es continua, luego $\Phi(K) = \left\{d^2f(a)\left(\frac{x-a}{\|x-a\|}, \frac{x-a}{\|x-a\|}\right)\right\}_{x\in\mathbb{R}^m}$ es compacto y, por ser además $d^2f(a)$ definida positiva, existe $M > 0$ tal que $\Phi(K) \geq M$.
Como $\frac{2o(\|x-a\|^2)}{\|x-a\|^2} \to 0$ cuando $x \to a$, existe $\mathcal{U} \in \mathcal{E}(a)$ tal que $\forall x \in \mathcal{U}, \left|\frac{2o(\|x-a\|)^2}{\|x-a\|^2}\right| < M$, luego si $x \in \mathcal{U}\backslash\{a\}, d^2f(a)\left(\frac{x-a}{\|x-a\|}, \frac{x-a}{\|x-a\|}\right) + \frac{2o(\|x-a\|^2)}{\|x-a\|^2} > M - M = 0$ y $f(x) > f(a) + \frac{1}{2}\|x-a\|^2 \cdot 0 = f(a)$.

$2 \implies 3]$ Obvio.

$3 \implies 4]$ Fijamos $u \in \mathbb{R}^m$ y definimos $\varphi(t) := a + tu$ como la función $\varphi : \mathbb{R} \to \mathbb{R}^m$ que parametriza la recta $a+ < \vec{u} >$. Sea $\mathcal{U} \in \mathcal{E}(a)$ con $f(a) \leq f(x) \forall x \in \mathcal{U}$, si restringimos φ a $\varphi^{-1}(\mathcal{U})$, un entorno de 0 en $\mathbb{R}$, entonces $f \circ \varphi$ alcanza un mínimo en 0, pues $(f \circ \varphi)(0) = f(\varphi(0)) = f(a) \leq f(\varphi(t)) \forall t \in \varphi^{-1}(\mathcal{U})$, y tenemos que $f \circ \varphi$ es de clase $\mathcal{C}^2$ y semidefinida positiva. Por la regla de la cadena, al ser φ y f diferenciables,

$$d(f \circ \varphi)(t) = df(\varphi(t)) \circ d\varphi(t) \equiv$$

$$\equiv \begin{pmatrix} \frac{\partial f}{\partial x_1}(a+tu) & \cdots & \frac{\partial f}{\partial x_m}(a+tu) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} = \sum_i \frac{\partial f}{\partial x_i}(a+tu)u_i$$

Entonces $d^2(f \circ \varphi)(t) \equiv \frac{d}{dt} \sum_i \frac{\partial f}{\partial x_i}(a+tu)u_i = \sum_i u_i \frac{d}{dt}\left(\frac{\partial f}{\partial x_i}(a+tu)\right)$. Como $\frac{\partial f}{\partial x_i}$ es diferenciable al ser f de clase $\mathcal{C}^2$,

$$\frac{d}{dt}\left(\frac{\partial f}{\partial x_i}(a+tu)\right) = d\left(\frac{\partial f}{\partial x_i} \circ \varphi\right)(t) = d\frac{\partial f}{\partial x_i}(\varphi(t)) \circ d\varphi(t) \equiv$$

$$\equiv \begin{pmatrix} \frac{\partial^2 f}{\partial_i \partial_1}(a+tu) & \cdots & \frac{\partial^2 f}{\partial_i \partial_m}(a+tu) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} = \sum_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a+tu)u_j$$

Sustituyendo, $d^2(f \circ \varphi)(t) \equiv \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(a+tu)u_i u_j = d^2f(a+tu)(u,u)$. Pero al ser $f \circ \varphi$ una función real de una variable real dos veces derivable con su mínimo en 0, sustituyendo $0 \leq (f \circ \varphi)''(0) = d^2f(a)(u,u)$, y como esto se cumple para todo $u \in \mathbb{R}^m$, queda probado que $d^2f(a)$ es semidefinida positiva.

Capítulo 4

Regiones de $\mathbb{R}^n$

Podemos describir una región de $\mathbb{R}^n$:

1. De forma **implícita**, como el conjunto de puntos que cumplen $f(x) = 0$ para cierta función $f : \mathcal{U} \to \mathbb{R}^k$, siendo $\mathcal{U} \subseteq \mathbb{R}^n$ un abierto. La región $A = \{(x_1, \dots, x_n) \in \mathcal{U} \mid f(x_1, \dots, x_n) = 0\}$ está **descrita implícitamente de forma $\mathcal{C}^1$-regular** si f es de clase $\mathcal{C}^1$ y $\forall p \in A, \operatorname{rg}(df(p)) = k$ (el rango de la diferencial es k).

2. De forma **paramétrica**, como la imagen de una función $\varphi : \mathcal{U} \to \mathbb{R}^n$, siendo $\mathcal{U} \subseteq \mathbb{R}^m$ un abierto. La **parametrización** es **$\mathcal{C}^1$-regular** si φ es de clase $\mathcal{C}^1$ y $\forall p \in \mathcal{U}, \operatorname{rg}(d\varphi(p)) = m$.

El **teorema de la función implícita**[1] afirma que, para $A \subseteq \mathbb{R}^n$ y $p \in A$, existe un $\mathcal{U} \in \mathcal{E}(p)$ tal que $\mathcal{U} \cap A$ admite una presentación implícita $\mathcal{C}^1$-regular si y sólo si existe $\mathcal{U}' \in \mathcal{E}(p)$ tal que $\mathcal{U} \cap A$ admite una parametrización $\mathcal{C}^1$-regular.

Sean pues $\mathcal{U} \subseteq \mathbb{R}^m \xrightarrow{\varphi} \mathcal{V} \subseteq \mathbb{R}^n \xrightarrow{f} \mathcal{W} \subseteq \mathbb{R}^k$ la parametrización y la forma implícita de A y $q \in \mathcal{U}$ tal que $\varphi(q) = p \in \mathcal{V}$, se tiene que $\operatorname{Im}(d\varphi(q)) = \ker(df(p))$. En efecto, como f es constante en A por ser $f(A) = \{0\}$, $f \circ \varphi$ también lo es, luego $0 = d(f \circ \varphi)(q) = df(p) \circ d\varphi(q)$ y entonces $\operatorname{Im}(d\varphi(q)) \subseteq \ker(df(p))$, pero como ambos subespacios tienen la misma dirección, se tiene la igualdad. Esto significa además que este espacio no depende de φ o f, y en esta situación llamamos **espacio tangente** al compacto A en el punto p al espacio afín que pasa por p y tiene por dirección $\operatorname{Im}(d\varphi(q)) = \ker(df(p))$.

Llamamos **gradiente** en $a \in \mathcal{U}$ de una función $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ diferenciable en a al vector $\nabla f(a) := \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right) \in \mathbb{R}^n$, la matriz de la diferencial de f en a expresada como vector. Para encontrar los extremos relativos de una función $f : D \to \mathbb{R}^n$ sobre un subconjunto $D \subseteq \mathbb{R}^n$ no abierto:

- Si D está dado en forma paramétrica como $\varphi(\mathcal{U})$, donde $\mathcal{U}$ es un abierto de $\mathbb{R}^m$ y $\varphi : \mathcal{U} \to \mathbb{R}^n$ es diferenciable, buscamos los extremos relativos de $f \circ \varphi$ en $\mathcal{U}$, teniendo en cuenta que $f \circ \varphi$ tiene máximo absoluto en a si y sólo si f tiene un máximo absoluto en $\varphi(a)$. Si además φ es continua, un máximo relativo de f en $\varphi(a)$ implica uno de $f \circ \varphi$ en a, y si $\varphi : (\mathcal{U}, \mathcal{T}_u|_{\mathcal{U}}) \to (\varphi(\mathcal{U}), \mathcal{T}_u|_{\varphi(\mathcal{U})})$ es abierta, un máximo relativo de $f \circ \varphi$ en a es uno de f en $\varphi(a)$.

[1] Esto corresponde a FVV3, pero lo estudiamos por su utilidad práctica.

- Si D está dado en forma implícita como $\{x \in \mathcal{U} \mid g(x) = 0\}$, donde $\mathcal{U}$ es un abierto de $\mathbb{R}^n$ y $g : \mathcal{U} \to \mathbb{R}^k$ es de clase $\mathcal{C}^1$, aplicamos el **teorema de los multiplicadores de Lagrange**, que afirma que si $f : \mathcal{U} \to \mathbb{R}$ es diferenciable, alcanza en un punto $a \in \mathcal{U}$ un extremo relativo y $\operatorname{rg}(dg(a)) = k$, entonces $\nabla f(a) \in \operatorname{span}(\nabla g_1(a), \dots, \nabla g_k(a)) :=< \nabla g_1(a), \dots, \nabla g_k(a) >$ (el espacio generado por los vectores). **Demostración:** Por el teorema de la función implícita, existen $\mathcal{V} \subseteq \mathbb{R}^{n-k}$ abierto, $\mathcal{W} \in \mathcal{E}(a)$ y $\varphi : \mathcal{V} \to \mathcal{W}$ de clase $\mathcal{C}^1$ con $\operatorname{rg}(d\varphi(a)) = n - k$ tales que $D \cap \mathcal{W} = \varphi(\mathcal{V})$, y si a es extremo relativo de f en D, por la continuidad de φ, el punto $b \in \mathcal{V}$ con $\varphi(b) = a$ es extremo relativo de $f \circ \varphi$ en $\mathcal{V}$, luego $d(f \circ \varphi)(b) = 0 = df(a) \circ \varphi(b)$ y entonces $\operatorname{Im}(d\varphi(b)) = \ker(dg(a)) = \ker(dg_1(a), \dots, dg_k(a)) \subseteq \ker(df(a))$ y por tanto $\bigcap_{i=1}^k \ker(dg_i(a)) \subseteq \ker(df(a))$, que por un misterioso lema de álgebra equivale a que $df(a) \in \operatorname{span}(dg_1(a), \dots, dg_k(a))$.