Lo esencial sobre funciones de clase C^k, polinomios de Taylor, extremos, y extremos condicionados.

Daniel Azagra Rueda

8 de diciembre de 2018

1. Funciones de clase C^k

Definición 1.1. Sean U un abierto de $\mathbb{R}^n$, y $k \in \mathbb{N}$. Diremos que una función $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ es de clase C^k, y escribiremos $f \in C^k(U)$, si todas sus derivadas parciales de orden k existen y son continuas en U. Diremos que $g : U \to \mathbb{R}^m$ es de clase C^k, y escribiremos $g \in C^k(U, \mathbb{R}^m)$, si cada función componente de g es de clase C^k.

En particular, $C^{k+1}(U, \mathbb{R}^m) \subset C^k(U, \mathbb{R}^m)$, y toda función de clase C^k es diferenciable.

Diremos que f es de clase C^∞, y escribiremos $f \in C^\infty(U)$, si f es de clase C^k para cada $k \in \mathbb{N}$.

Problema 1.1. Demostrar que la suma, el producto y la composición de funciones de clase C^k es de clase C^k.

Si una función es de clase C^2 entonces sus derivadas parciales segundas tienen una importante propiedad de simetría.

Teorema 1.1 (Schwarz). *Sea $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ tal que las derivadas parciales $\partial^2 f/\partial x_i \partial x_j$ y $\partial^2 f/\partial x_j \partial x_i$ existen y son continuas en U. Entonces*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Demostración. Podemos suponer $m = 1$. Además, dejando quietas todas las variables menos dos y aplicando la definición de derivada parcial, es obvio que podemos también suponer $n = 2$. Así, queremos ver que

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{\partial^2 f}{\partial y \partial x}(x, y)$$

para cada $(x, y) \in U$. Fijemos $(x, y) \in U$, y para todos $h, k \in \mathbb{R}$ suficientemente pequeños, consideremos la cantidad

$$S_{h,k} = [f(x + h, y + k) - f(x, y + k)] - [f(x + h, y) - f(x, y)].$$

Definamos para cada k la función g_k por

$$g_k(u) = f(u, y + k) - f(u, y),$$

y obsérvese que

$$S_{h,k} = g_k(x + h) - g_k(x).$$

Aplicando el teorema del valor medio, podemos escribir entonces

$$S_{h,k} = g_k'(c_{k,h})h$$

para cierto $c_{k,h}$ entre x y $x+h$. Es decir,

$$S_{h,k} = \left(\frac{\partial f}{\partial x}(c_{k,h}, y+k) - \frac{\partial f}{\partial x}(c_{k,h}, y)\right) h = \frac{\partial^2 f}{\partial y \partial x}(c_{k,h}, d_{k,h}) hk \tag{1}$$

para cierto $d_{k,h}$ entre y e $y+k$, si aplicamos otra vez el teorema del valor medio a la diferencia de las primeras derivadas parciales.

Puesto que la expresión $S_{h,k}$ es simétrica en h, k y x, y, intercambiando los dos términos del medio en la expresión que define $S_{h,k}$, podemos deducir de manera análoga que

$$S_{h,k} = \frac{\partial^2 f}{\partial x \partial y}(\tilde{c}_{k,h}, \tilde{d}_{k,h}) hk \tag{2}$$

para ciertos $\tilde{c}_{k,h}, \tilde{d}_{k,h}$ entre x y $x+h$ e y e $y+k$. Igualando (1) y (2), cancelando hk, y haciendo $(h,k) \to (0,0)$ y usando la continuidad de las derivadas parciales segundas, llegamos al resultado buscado. □

Sin embargo, la mera existencia de las derivadas parciales cruzadas de segundo orden, no garantiza su igualdad, como prueba el ejemplo siguiente.

Problema 1.2. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{si } (x,y) \neq (0,0), \\ 0 & \text{si } (x,y) = (0,0). \end{cases}$$

Calcular las derivadas parciales de primer y segundo orden de f, y comprobar que

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1 \neq -1 = \frac{\partial^2 f}{\partial y \partial x}(0,0).$$

Definición 1.2. Si $f \in C^2(U)$, definiremos su diferencial de orden 2 en un punto $a \in U$ como la forma bilineal

$$D^2 f(a)(u,v) = \sum_{i,j=1}^{n} \frac{\partial f}{\partial x_i \partial x_j}(a) u_i v_j.$$

A la matriz cuyas entradas son las derivadas parciales de orden 2 de f en un punto a, es decir,

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(a) \\ \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{pmatrix},$$

se le llama matriz hessiana de f en a. Si denotamos esta matriz $H_{f,a}$, nótese que se tiene

$$D^2 f(a)(u,v) = \langle v, H_{f,a} u \rangle.$$

Por el teorema de Schwarz, la forma bilineal $D^2 f(a)$ es simétrica.

Definición 1.3. Más en general, si $f \in C^k(U)$ con $k \geq 2$, y $a \in U$, definiremos $D^k f(a)$ como la forma multilineal simétrica de orden k

$$D^k f(a)(u^1, ..., u^k) = \sum_{i_1, \dots, i_k = 1}^{n} \frac{\partial^k f}{\partial x_{i_1} ... \partial x_{i_k}}(a) u^1_{i_1} ... u^k_{i_k}.$$

A veces usaremos una forma algo más concisa de escribir las derivadas parciales:

$$D_{i_1, \dots, i_k} f(a) := \frac{\partial^k f}{\partial x_{i_1} ... \partial x_{i_k}}(a).$$

Con esta notación tenemos

$$D^k f(a)(u^1, ..., u^k) = \sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} D_{i_1,...,i_k} f(a) u^1_{i_1} ... u^k_{i_k}.$$

En el caso en que tengamos $u^1 = u^2 = ... = u^k$, denotaremos $D^k f(a)(u)^k$ en lugar de $D^k f(a)(u, ..., u)$.

Es claro que la forma k-lineal $D^k f(x)$ queda unívocamente determinada por las derivadas parciales $D_{i_1,...,i_k} f(x)$, $i_1, ..., i_k \in \{1, ..., n\}$, y la aplicación $U \ni x \mapsto Df(x)$ puede verse por tanto como una función definida en U y que toma valores en $\mathbb{R}^{n^k}$. Podemos por tanto investigar su diferenciabilidad en un punto $a \in U$. Como una función es diferenciable si y sólo si lo son sus funciones componentes, esta función $D^k f$ será diferenciable en a si y sólo si lo son cada una de las derivadas parciales $D_{i_1,...,i_k} f$. En tal caso existirán todas las derivadas parciales de cada una de estas funciones, es decir, existirán $D_{i_1,...,i_k,i_{k+1}} f(a)$, para todos los $i_1, ..., i_k, i_{k+1} \in \{1, ..., n\}$. El siguiente teorema, análogo al de Schwarz, nos garantiza que en esta situación se tiene que $D_{i_1,...,i_k,i_{k+1}} f(a) = D_{j_1,...,j_k,j_{k+1}} f(a)$ siempre que el conjunto de índices $\{j_1, ..., j_k, j_{k+1}\}$ sea una permutación de $\{i_1, ..., i_k, i_{k+1}\}$. Por tanto podremos identificar la derivada de $D^k f(x)$ en a con la forma $(k+1)$-lineal simétrica definida por

$$D^k f(a)(u^1, ..., u^k, u^{k+1}) = \sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} \sum_{i_{k+1}=1}^{n} D_{i_1,...,i_k,i_{k+1}} f(a) u^1_{i_1} ... u^k_{i_k} u^{k+1}_{i_{k+1}}.$$

En particular, cuando f es de clase C^{k+1} en U, $D^k f$ siempre es diferenciable en U, y su diferencial es $D^{k+1} f$.

Teorema 1.2. *Sea $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ una función diferenciable en U, y supongamos que*

$$\frac{\partial f}{\partial x_i}, \frac{\partial f}{\partial x_j} : U \to \mathbb{R}$$

son ambas diferenciables en $a \in U$. Entonces

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a).$$

En particular, si f es dos veces diferenciable en a, la matriz de su diferencial segunda $D^2 f(a)$ es simétrica.

Demostración. Es parecida a la del teorema de Schwarz 1.1, y como en este caso podemos suponer $n = 2$.

Fijemos $(a, b) \in U$, y para todo $h \in \mathbb{R}$ suficientemente pequeño, consideremos la función

$$F(h) = f(a+h, b+h) - f(a+h, b) - f(a, b+h) + f(a, b)].$$

Si ponemos $\varphi(x) = f(x, b+h) - f(x, b)$, entonces $F(h) = \varphi(a+h) - \varphi(a)$, y aplicando el teorema del valor medio obtenemos

$$F(h) = \varphi'(t_h)h = [\frac{\partial f}{\partial x}(t_h, b+h) - \frac{\partial f}{\partial x}(t_h, b)]h,$$

para cierto t_h entre a y $a+h$. Como $\partial f/\partial x$ es diferenciable en (a, b), podemos escribir

$$\frac{\partial f}{\partial x}(t_h, b+h) = \frac{\partial f}{\partial x}(a, b) + \frac{\partial^2 f}{\partial x^2}(a, b)(t_h - a) + \frac{\partial^2 f}{\partial y \partial x}(a, b)h + \epsilon_1(h)|h|,$$

y también

$$\frac{\partial f}{\partial x}(t_h, b) = \frac{\partial f}{\partial x}(a, b) + \frac{\partial^2 f}{\partial x^2}(a, b)(t_h - a) + \epsilon_2(h)|h|,$$

donde ϵ_1, ϵ_2 son funciones de h tales que $\lim_{h\to 0} \epsilon_i(h) = 0$. Por tanto

$$F(h) = \frac{\partial^2 f}{\partial y \partial x}(a,b)h^2 + \epsilon(h)|h|^2, \tag{1}$$

donde $\epsilon(h) = \epsilon_1(h) - \epsilon_2(h)$ y en particular $\lim_{h\to 0} \epsilon(h) = 0$.

Por otra parte, si ponemos $\psi(y) = f(a+h, y) - f(a, y)$, se tiene $F(h) = \psi(b+h) - \psi(b)$, y una repetición del argumento anterior nos da que

$$F(h) = \frac{\partial^2 f}{\partial x \partial y}(a,b)h^2 + \overline{\epsilon}(h)|h|^2, \tag{2}$$

donde $\lim_{h\to 0} \overline{\epsilon}(h) = 0$. Juntando (1) y (2) obtenemos que

$$\frac{\partial^2 f}{\partial y \partial x}(a,b) = \lim_{h\to 0} \frac{F(h)}{h^2} = \frac{\partial^2 f}{\partial x \partial y}(a,b),$$

que es lo que pretendíamos demostrar. □

2. Polinomios de Taylor

Definición 2.1. *Sean $U \subseteq \mathbb{R}^n$ abierto, $f : U \to \mathbb{R}$ una función de clase C^m. Se define el polinomio de Taylor de orden m de f en $a \in U$ como*

$$P^k_{f,a}(h) = f(a) + Df(a)(h) + \frac{1}{2!}D^2 f(a)(h)^2 + \frac{1}{3!}D^3 f(a)(h)^3 + \dots + \frac{1}{m!}D^m f(a)(h)^m.$$

Ejemplo 2.1. Calcular el polinomio de Taylor de orden 3 de la función $f(x,y) = x\cos y$ en el punto $a = (1,0)$.

A continuación vemos cómo la versión general del teorema de Taylor puede deducirse fácilmente de la versión ya conocida para funciones de variable real.

Teorema 2.1 (fórmula de Taylor)**.** *Sean U abierto de $\mathbb{R}^n$, $f : U \to \mathbb{R}$ una función $m+1$ veces diferenciable en U, con $m \geq 1$, $a \in U$, $h \in \mathbb{R}^n$ con $[a, a+h] \subset U$. Entonces existe $\xi = \xi_h \in [a, a+h]$ tal que*

$$\begin{aligned} f(a+h) = \quad & f(a) + Df(a)(h) + \tfrac{1}{2!}D^2 f(a)(h)^2 + \dots + \tfrac{1}{m!}D^m f(a)(h)^m + \\ & + \tfrac{1}{(m+1)!}D^{m+1} f(\xi_h)(h)^{m+1} = \\ & = P^m_{f,a}(h) + \tfrac{1}{(m+1)!}D^{m+1} f(\xi_h)(h)^{m+1}. \end{aligned}$$

Por otra parte, si $k \geq 2$, suponiendo únicamente que f es $k-1$ veces diferenciable en el abierto U, y que $D^{k-1}f$ es diferenciable en un punto $x \in U$, se tiene que

$$\lim_{h\to 0} \frac{f(x+h) - P^k_{f,x}(h)}{\|h\|^k} = 0.$$

Demostración. Definamos

$$g(t) = f(a+th),$$

para $t \in [0,1]$. Por el teorema de Taylor para funciones de una variable real, existe $\zeta \in [0,1]$ tal que

$$g(1) = g(0) + \sum_{k=1}^{m} \frac{1}{k!} g^{(k)}(0) + \frac{1}{(m+1)!} g^{(m+1)}(\zeta).$$

Nótese que $g(1) = f(a+h)$, y $g(0) = f(a)$. Además,

$$g^{(k)}(t) = D^k f(a+th)(h)^k.$$

En efecto, por la regla de la cadena, $g'(t) = \sum_{i=1}^n D_i f(a+th)h_i$; aplicando nuevamente la regla de la cadena, $g''(t) = \sum_{i=1}^n \frac{d}{dt} D_i f(a+th)h_i = \sum_{i=1}^n \sum_{j=1}^n D_{ji} f(a+th)h_j h_i$, y así sucesivamente se ve que $g^{(k)}(t) = \sum_{i_1,...,i_k=1}^n D_{i_1,...,i_k} f(a+th)h_{i_1}...h_{i_k}$.

Por tanto, sustituyendo,

$$f(a+h) = f(a) + \sum_{k=1}^m \frac{1}{k!} D^k f(a)(h) + \frac{1}{(m+1)!} D^{m+1} f(a+\zeta h)(h),$$

lo que prueba la primera parte del enunciado si ponemos $\xi_h = a + \zeta h$.

Para ver la segunda parte del enunciado, para cada dirección $v \in \mathbb{R}^n$, $\|v\| = 1$, y cada $t \in (0,r)$, donde r es tal que $B(x,r) \subseteq U$, aplicamos el teorema del valor medio de Cauchy para encontrar un número $\alpha_{t,v} \in (0,t)$ tal que

$$\frac{f(x+tv) - f(x) - tDf(x)(v) - ... - \frac{1}{k!} t^k D^k f(x)(v)^k}{t^k} =$$
$$\frac{Df(x+\alpha_{t,v} v)(v) - Df(x)(v) - \alpha_{t,v} D^2 f(x)(v)^2 - ... - \frac{1}{(k-1)!} \alpha_{t,v}^{k-1} D^k f(x)(v)^k}{k\alpha_{t,v}^{k-1}}.$$

Volviendo a aplicar el teorema del valor medio de Cauchy obtenemos un número $\beta_{t,v} \in (0, \alpha_{t,v})$ tal que

$$\frac{Df(x+\alpha_{t,v} v)(v) - Df(x)(v) - \alpha_{t,v} D^2 f(x)(v)^2 - ... - \frac{1}{(k-1)!} \alpha_{t,v}^{k-1} D^k f(x)(v)^k}{k\alpha_{t,v}^{k-1}} =$$
$$\frac{D^2 f(x+\beta_{t,v} v)(v) - D^2 f(x)(v)^2 - \beta_{t,v} D^3 f(x)(v)^3 - ... - \frac{1}{(k-2)!} \beta_{t,v}^{k-2} D^k f(x)(v)^k}{k(k-1)\beta_{t,v}^{k-2}}.$$

Aplicando reiteradamente el teorema del valor medio de Cauchy hasta $k-3$ veces más, obtenemos un número $s_{t,v} \in (0,t)$ tal que

$$\frac{f(x+tv) - f(x) - tDf(x)(v) - ... - \frac{1}{k!} t^k D^k f(x)(v)^k}{t^k} = \tag{2.1}$$
$$\frac{D^{k-1} f(x+s_{t,v} v)(v)^{k-1} - D^{k-1} f(x)(v)^{k-1} - s_{t,v} D^k f(x)(v)^k}{k! s_{t,v}}. \tag{2.2}$$

Ahora bien, puesto que $D^{k-1} f$ es diferenciable en x, y su diferencial es $D^k f(x)$, se tiene que

$$\lim_{s \to 0^+} \frac{D^{k-1} f(x+sv)(v)^{k-1} - D^{k-1} f(x)(v)^{k-1} - sD^k f(x)(v)^k}{k! s} = 0,$$

uniformemente en $v \in \mathbb{S}^{n-1}$ (la esfera unidad de $\mathbb{R}^n$). Es decir, dado $\varepsilon > 0$ existe $\delta > 0$ tal que si $0 < s < \delta$ entonces

$$\left| \frac{D^{k-1} f(x+sv)(v) - D^{k-1} f(x)(v) - sD^k f(x)(v)}{k! s} \right| \leq \varepsilon$$

para todo $v \in \mathbb{S}^n$. Luego, si $0 < t < \delta$, como $s_{t,v} \in (0,t) \subset (0,\delta)$ para todo $v \in \mathbb{S}^n$, deducimos de aquí y de (2.1) que si $0 < t < \delta$ entonces

$$\left| \frac{f(x+tv) - f(x) - tDf(x)(v) - ... - \frac{1}{k!} t^k D^k f(x)(v)^k}{t^k} \right| \leq \varepsilon$$

para todo $v \in \mathbb{S}^n$. Esto significa que

$$\lim_{t\to 0^+} \frac{f(x+tv)-f(x)-tDf(x)(v)-...-\frac{1}{k!}t^kD^kf(x)(v)^k}{t^k}=0$$

uniformemente en $v \in \mathbb{S}^{n-1}$, lo que equivale a decir que

$$\lim_{h\to 0} \frac{f(x+h)-f(x)-Df(x)(h)-...-\frac{1}{k!}D^kf(x)(h)^k}{|h|^k}=0,$$

como queríamos comprobar. □

De manera análoga, usando la versión del teorema de Taylor con resto en forma integral para funciones de una variable, se demuestra el siguiente

Teorema 2.2 (fórmula de Taylor con resto integral)**.** *Sean U abierto de $\mathbb{R}^n$, $f : U \to \mathbb{R}$ una función de clase m en U, con $m \geq 1$, $a \in U$, $h \in \mathbb{R}^n$ con $[a, a+h] \subset U$. Entonces*

$$f(a+h)-\left(f(a)+Df(a)(h)+\frac{1}{2!}D^2f(a)(h)^2+...+\frac{1}{m!}D^mf(a)(h)^m\right)=$$
$$\int_0^1 \frac{(1-t)^{m-1}}{(m-1)!}\left(D^mf(a+th)(h)^m-D^mf(a)(h)^m\right)dt.$$

Observación 2.3. A la cantidad

$$R^m_{f,a}(h) := f(a+h)-P^m_{f,a}(h)$$

se le suele llamar resto de Taylor. Así la segunda parte del teorema de Taylor (o la anterior fórmula integral del resto de Taylor, si suponemos que además f es de clase C^m en un entorno de a) nos dice que

$$\lim_{h\to 0}\frac{R^m_{f,a}(h)}{\|h\|^m}=0.$$

Supongamos que $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ es una función de clase C^∞. En cada punto $a \in U$ podemos formar una serie infinita con las derivadas de f,

$$\sum_{k=0}^{\infty}\frac{1}{k!}D^kf(a)(x-a)$$

(donde $D^0 = f$). A esta serie se le llama la *serie de Taylor* de f en el punto a. Supongamos también que esta serie es convergente para x en un entorno de a. Es natural preguntarse si se tendrá

$$f(x)=\sum_{k=0}^{\infty}\frac{1}{k!}D^kf(a)(x-a)$$

al menos para todo x en un entorno suficientemente pequeño de a. Esto no es siempre así, como prueba el ejemplo siguiente.

Ejemplo 2.2. Sea $f : \mathbb{R} \to \mathbb{R}$ definida por

$$f(t)=\begin{cases} e^{-1/t^2} & \text{si } t\neq 0,\\ 0 & \text{si } t=0.\end{cases}.$$

No es difícil comprobar que $f \in C^\infty(\mathbb{R},\mathbb{R})$ y que $D^kf(0) = 0$ para todo $k \in \mathbb{N}$. Por tanto, la serie de Taylor de f en 0 es constantemente cero, y sin embargo para todo $\delta > 0$, se tiene $f(\delta) \neq 0$, luego la serie de Taylor de f no converge a la función f en nigún entorno de 0.

Definición 2.2. *Se dice que $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ es real-analítica en U si para cada $a \in U$ existe $r > 0$ tal que*

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} D^k f(a)(x-a)^k$$

para todo $x \in B(a,r)$. Es decir, f es real analítica si el resto de Taylor $R^k_{f,a}$ converge a cero cuando $k \to \infty$ en un entorno de cada $a \in U$.

En este curso no estudiaremos en profundidad las funciones real-analíticas; lo natural es hacerlo en el marco de la teoría de funciones de variable compleja. Baste con decir que los polinomios, las funciones trigonométricas, y las funciones exponencial y logaritmo son real-analíticas, y que puede demostrarse que la suma, producto, cociente (donde esté bien definido) y composición de funciones real-analíticas son real-analíticas.

Problemas

Problema 2.1. Calcúlese el Polinomio de Taylor de grado menor o igual que 3 para las funciones:

$$f(x,y) = e^{(x-1)^2}\cos y \text{ en } (1,0), \quad g(x,y) = x\sin y + y \sin x \text{ en } (0,0),$$
$$h(x,y) = xy^2 \text{ en } (0,0), \qquad \varphi(x,y) = \log(x+y) \text{ en } (1,1),$$
$$\psi(x,y) = \frac{1}{x^2+y^2+1} \text{ en } (0,0), \qquad \Phi(x,y) = \tfrac{y^2}{x^3} \text{ en } (1,-1).$$

Problema 2.2. Escríbase el polinomio $x^3 + y^2(1+x)$ en potencias de $(x-1)$ e $(y-2)$.

Problema 2.3. Demuéstrese que $f(x,y) = y^x$ es diferenciable para $y > 0$. Encuéntrese un polinomio $P(x,y)$ en dos variables, de grado menor o igual que 3, tal que

$$\lim_{(x,y)\to(0,0)} \frac{f(x,1+y) - P(x,y)}{(x^2+y^2)^{3/2}} = 0.$$

3. Extremos locales y puntos críticos

Como es bien conocido, para encontrar los máximos y mínimos locales de una función real definida sobre un intervalo de $\mathbb{R}$, la técnica más socorrida es hallar los puntos donde su derivada se anula (los puntos críticos), y ver después cuál es el signo de la derivada segunda sobre esos puntos críticos. Si se quiere después hallar los máximos y mínimos absolutos de la función en el intervalo, se comparan los valores de f en los puntos críticos con los que toma en los extremos del intervalo. Como veremos en esta sección, el mismo método, esencialmente, sirve para funciones reales de varias variables, con algunas diferencias importantes. Primera, que la segunda derivada ya no es un número, sino una forma bilineal simétrica. En este caso el *signo* de la derivada será sustituido por ser definida positiva o definida negativa. Segunda, que la frontera del conjunto abierto donde esté definida la función ya no va a ser un conjunto tan simple como un par de puntos: en general habrá infinitos puntos en dicha frontera, y la tarea de comparar uno por uno los valores de f en estos puntos es obviamente impracticable. Será por tanto necesario desarrollar nuevas técnicas que permitan hallar los máximos y mínimos de una función diferenciable restringida a la frontera de un abierto. Esto, dicho así en toda generalidad, es un objetivo excesivamente ambicioso, pero si la frontera del abierto es un conjunto suficientemente regular (por ejemplo, una hipersuperficie diferenciable) entonces el problema tendrá solución satisfactoria.

Definición 3.1. Se dice que $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ tiene un máximo local (o relativo) en un punto $a \in A$ si existe $r > 0$ tal que

$$f(x) \leq f(a) \text{ para todo } x \in A \cap B(x,r).$$

Se dice que f tiene un mínimo local en a si existe $r > 0$ tal que

$$f(a) \leq f(x) \text{ para todo } x \in A \cap B(x, r).$$

En cualquiera de los casos se dirá que f tiene un extremo local en a, y que es estricto si la desigualdad es estricta cuando $x \neq a$.

Definición 3.2. Sea $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ una función diferenciable en $a \in U$. Se dice que a es un *punto crítico* de f si $Df(a) = 0$.

Como ocurre con las funciones de una variable, todos los extremos locales de las funciones diferenciables son puntos críticos.

Proposición 3.1. *Sean U abierto de $\mathbb{R}^n$, y $f : U \to \mathbb{R}$ diferenciable en U. Supongamos que f tiene un extremo local en $a \in U$. Entonces*

$$Df(a) = 0,$$

es decir a es un punto crítico de f.

Demostración. Sea $v \in \mathbb{R}^n$ cualquiera con $\|v\| = 1$. Definamos $g_v : (-r, r) \to \mathbb{R}$ por

$$g_v(t) = f(a + tv),$$

donde $r > 0$ es tal que $B(a, r) \subseteq U$. Como f tiene un extremo local en a, es evidente que g_v también tiene un extremo local (del mismo tipo) en $t = 0$. Entonces, por el resultado ya conocido para funciones de variable real, se tiene que

$$0 = g'(0) = Df(a)(v).$$

Esto prueba que $Df(a)(v) = 0$ para todo v con $\|v\| = 1$, y por tanto, como $Df(a)$ es lineal, $Df(a) = 0$. □

El recíproco no es cierto, es decir, no todo punto crítico de una función diferenciable es un extremo local. Por ejemplo, la función $f : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$f(x, y) = xy$$

tiene un punto crítico en $(0, 0)$, que no es ni máximo ni mínimo (téngase en cuenta que $f(t, t) = t^2$, y $f(t, -t) = -t^2$). La gráfica de f es la típica silla de montar, y de $(0, 0)$ se dice que es un *punto de silla* de f.

Para determinar si un punto crítico es un máximo o un mínimo local, o ninguna de las dos cosas, es suficiente, en muchos casos, un análisis de la matriz de la segunda derivada.

Definición 3.3. Sea $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ una función de clase C^2. Se llama hessiana de f en a a la segunda derivada de f en a, y matriz hessiana a la matriz de su forma cuadrática asociada $Q(h) = D^2 f(a)(h)^2$, es decir,

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(a) \\ \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{pmatrix}.$$

Recordemos que, si Q es una forma cuadrática en $\mathbb{R}^n$, se dice que:

1. Q es definida positiva si $Q(h) > 0$ para todo $h \neq 0$;

2. Q es semidefinida positiva si $Q(h) \geq 0$ para todo $h \in \mathbb{R}^n$;

3. Q es definida negativa si $Q(h) < 0$ para todo $h \neq 0$;

4. Q es semidefinida negativa si $Q(h) \leq 0$ para todo $h \in \mathbb{R}^n$;

5. Q es indefinida si no es ni semidefinida positiva ni semidefinida negativa, es decir, si existen $h_1, h_2 \in \mathbb{R}^n$ tales que $Q(h_1) < 0 < Q(h_2)$.

El teorema de Taylor nos permite dar condiciones suficientes sobre la segunda derivada de una función en un punto crítico para asegurar que se trata de un mínimo o un máximo local, o decidir que no es ni lo uno ni lo otro.

Teorema 3.2. *Sean U un abierto de $\mathbb{R}^n$, y $f \in C^2(U, \mathbb{R})$. Supongamos que $Df(a) = 0$, es decir, $a \in U$ es un punto crítico de f. Entonces:*

1. *Si $D^2 f(a)$ es definida positiva, f tiene un mínimo local estricto en a.*

2. *Si $D^2 f(a)$ es definida negativa, f tiene un máximo local estricto en a.*

3. *Si $D^2 f(a)$ es indefinida, f tiene un punto de silla en a (es decir, no tiene mi máximo ni mínimo local en este punto crítico).*

   Además, se tiene un recíproco parcial (no total; ver los ejercicios) de lo anterior:

4. *Si f tiene un mínimo local en a entonces $D^2 f(a)$ es semidefinida positiva.*

5. *Si f tiene un máximo local en a entonces $D^2 f(a)$ es semidefinida negativa.*

Demostración. (1) Por el teorema de Taylor, y puesto que $Df(a) = 0$, sabemos que

$$f(a+h) = f(a) + \frac{1}{2} D^2 f(a)(h)^2 + R(h),$$

donde $\lim_{h \to 0} R(h)/\|h\|^2 = 0$. Sea

$$m := \inf\{D^2 f(a)(v)^2 : \|v\| = 1\}.$$

Como $D^2 f(a)(\cdot)$ es continua y la esfera unidad $S = \{h : \|h\| = 1\}$ es compacta, este ínfimo es en realidad un mínimo, y puesto que $D^2 f(a)$ es definida positiva, debe ser $m > 0$. Así, para todo $h \in \mathbb{R}^n$, $h \neq 0$, tenemos

$$D^2 f(a)\left(\frac{h}{\|h\|}\right)^2 \geq m > 0,$$

luego

$$D^2 f(a)(h)^2 \geq m\|h\|^2$$

para todo $h \in \mathbb{R}^n$. Sea $\varepsilon \in (0, m)$. Como $\lim_{h \to 0} R(h)/\|h\|^2 = 0$, existe $\delta > 0$ tal que si $\|h\| \leq \delta$ entonces $|R(h)| \leq \varepsilon \|h\|^2$, y por tanto

$$|f(a+h) - f(a) - D^2 f(a)(h)^2| \leq \varepsilon \|h\|^2,$$

luego

$$f(a+h) - f(a) \geq D^2 f(a)(h)^2 - \varepsilon\|h\|^2 \geq (m - \varepsilon)\|h\|^2$$

para todo h con $\|h\| \leq \delta$. Esto significa que

$$f(x) \geq f(a) \;\text{ si }\; x \in B(a, \delta),$$

y que la desigualdad es estricta cuando $x \neq a$, es decir, f tiene un mínimo local estricto en a.
(2): es análogo (o bien aplicar (1) a la función $-f$).
(4): Sea $v \in \mathbb{R}^n$ con $\|v\| = 1$. Por el teorema de Taylor, y teniendo en cuenta que f tiene un mínimo local en a, podemos escribir

$$0 \leq f(a+tv) - f(a) = \frac{1}{2} D^2 f(a)(tv)^2 + R(tv)$$

si t es suficientemente pequeño, donde $\lim_{t\to 0} R(tv)/\|tv\|^2 = 0$. Entonces, tomando límites,

$$0 \leq \lim_{t\to 0} \frac{f(a+tv)-f(a)}{t^2} = \lim_{t\to 0} \frac{1}{t^2}\left(\frac{1}{2}D^2 f(a)(tv)^2 + R(tv)\right) =$$
$$\lim_{t\to 0}\left(\frac{1}{2}D^2 f(a)(v)^2 + \frac{R(tv)}{\|tv\|^2}\right) = \frac{1}{2}D^2 f(a)(v)^2 + 0,$$

es decir, $D^2 f(a)(v)^2 \geq 0$ para todo v con $\|v\| = 1$, lo que significa que f es semidefinida positiva.
(5): análogo (o considerar $-f$).
(3): es consecuencia inmediata de (4) y (5). □

A continuación recordamos algunos criterios para decidir si una forma cuadrática simétrica (por ejemplo $D^2 f(a)$) es definida positiva, definida negativa, o indefinida.

Proposición 3.3. *Sea Q una forma cuadrática simétrica en $\mathbb{R}^n$. Entonces*

1. *Q es definida positiva si y sólo si todos sus autovalores son estrictamente positivos;*
2. *Q es definida negativa si y sólo si todos sus autovalores son estrictamente negativos;*
3. *Q es indefinida si y sólo si tiene autovalores positivos y negativos (estrictamente);*
4. *Q es semidefinida positiva si y sólo si todos sus autovalores son mayores o iguales que cero;*
5. *Q es semidefinida negativa si y sólo si todos sus autovalores son menores o iguales que cero.*

Recordemos también que los autovalores de Q pueden calcularse hallando las raíces de su polinomio característico,

$$P(\lambda) = \det(A - \lambda I),$$

donde A es la matriz de Q, y que todas las matrices correspondientes a formas cuadráticas simétricas tienen autovalores reales y son diagonalizables. Por tanto, para decidir si $D^2 f(a)$ es definida positiva o definida negativa, o indefinida, basta hallar sus autovalores y mirar el signo de éstos.

Si uno no desea calcular autovalores, el siguiente criterio puede ser igualmente útil, por lo menos para dimensiones bajas.

Proposición 3.4 (criterio de Silvester)**.** *Sea Q una forma cuadrática simétrica en $\mathbb{R}^n$, con matriz $A = (a_{ij})$. Consideremos los menores angulares de A,*

$$A_j = \det\begin{pmatrix} a_{11} & ... & a_{1j} \\ \vdots & ... & \vdots \\ a_{j1} & ... & a_{jj} \end{pmatrix}.$$

Entonces:

1. *Q es definida positiva si y sólo si $A_j > 0$ para todo $j = 1, ..., n$;*
2. *Q es definida negativa si y sólo si $(-1)^j A_j > 0$ para todo $j = 1, ..., n$;*
3. *Si Q es semidefinida positiva entonces $A_j \geq 0$ para todo $j = 1, ..., n$;*
4. *Si Q es semidefinida negativa entonces $(-1)^j A_j \geq 0$ para todo $j = 1, ..., n$.*

Si algún A_j con j par es negativo, entonces Q es indefinida.

4. Extremos globales y multiplicadores de Lagrange

Pasamos ahora a considerar brevemente la cuestión de cómo identificar los máximos y mínimos globales de una función en un conjunto no necesariamente abierto.

Lo primero de todo es asegurarse de que existen los extremos globales que se buscan. En general no tienen por qué existir si el conjunto no es compacto o la función no es continua, como ya sabemos. Incluso si la función es continua y acotada, el ínfimo y el supremo de su rango podrían no alcanzarse, como en el caso de la función $f:\mathbb{R}\to\mathbb{R}$, $f(x)=\arctan(x)$.

Por tanto, supondremos que nuestra función f está definida sobre un compacto K de $\mathbb{R}^n$ que sea la adherencia de un conjunto abierto acotado U sobre el cual la función f es diferenciable, y que además f es continua en K. En estas condiciones sabemos que f alcanza un máximo y un mínimo absolutos sobre K. La cuestión a la que debemos intentar responder ahora es cómo calcularlos y localizar los puntos donde se alcanzan. Cada caso concreto puede ser diferente, pero podemos dar algunas reglas generales. Lo primero que podemos hacer es intentar hallar los puntos críticos de f en U, tarea que en general puede ser muy difícil en sí misma, ya que puede ser imposible resolver explícitamente el sistema de ecuaciones $Df(x)=0$, aún suponiendo que tenga solución. Esto puede ocurrir perfectamente aún en el caso de las funciones más simples como los polinomios (es bien conocido que no hay fórmulas algebraicas para resolver ecuaciones polinómicas de grado igual o superior a cinco).

Seamos optimistas y supongamos que hemos sido capaces de resolver este sistema de ecuaciones $Df(x)=0$, y supongamos para empezar que tiene un número finito de soluciones, $a_1,...,a_k\in U$. Entre estos puntos críticos puede que se encuentre el máximo o el mínimo de f en K. Para decidir si esto es así, habrá que comparar los valores de f en los puntos a_i con los valores que toma f en la frontera $\partial K=\partial U$. El problema es que en dimensiones $n\geq 2$, ∂U consta siempre de una cantidad infinita de puntos, de modo que esto no puede hacerse de uno en uno como en el caso de funciones de una variable definidas en intervalos. Pero puede que hayamos tenido suerte y de alguna manera podamos saber lo que valen

$$\alpha=\text{ínf}\{f(x):x\in\partial U\}\ \text{ y }\ \beta=\sup\{f(x):x\in\partial U\}.$$

En este caso, el mayor de los valores $\{\alpha,\beta,f(a_1),...,f(a_k)\}$ corresponderá al máximo de f en K, y el menor de ellos al mínimo. El siguiente ejemplo ilustra esta situación.

Ejemplo 4.1. Hallar los máximos y mínimos de $f(x,y)=x^2+x+y^3-3y$ en el conjunto $K=\{(x,y)\in\mathbb{R}^2: x^2+x+y^3-3y\leq 7\}$.

En este ejemplo $\alpha=\beta=7$, y basta hallar los extremos relativos de f en el interior de K y comparar el valor de f en estos puntos con 7.

Un caso un poco más complicado que puede presentarse es que el conjunto de puntos críticos de f no sea finito. Parecería entonces que nuestra tarea está abocada al fracaso, pero en muchas situaciones sucede que puede identificarse una sucesión (finita o infinita) de conjuntos (infinitos) de puntos críticos sobre los cuales f es constante. Es decir, que a veces pueden encontrarse $C_k\subset K$, $k\in\mathbb{N}$, tales que $Df(x)=0$ si y sólo si $x\in\bigcup_{k=1}^{\infty}C_k$, y f es constante en C_k para cada k, pongamos $f(C_k)=\{c_k\}$. Entonces el mayor de los valores $\{\alpha,\beta\}\cup\{c_k:k\in\mathbb{N}\}$ corresponderá al máximo de f en K, y el menor de ellos al mínimo (¿por qué este conjunto, en general infinito, tiene un máximo y un mínimo?).

Ejemplo 4.2. Sea $F:\mathbb{R}^2\to\mathbb{R}$ definida por

$$f(x,y)=\begin{cases} e^{-1/(x^2+y^2)}\sin\frac{1}{x^2+y^2} & \text{si } (x,y)\neq(0,0),\\ 0 & \text{si } (x,y)=(0,0).\end{cases},$$

y sea $K=\{(x,y)\in\mathbb{R}^2: x^2+y^2\leq 5\}$. La función f es de clase C^∞ (por ser composición de la función $h(t)=e^{-1/t}\sin(1/t)$, $h(0)=0$, que es de clase C^∞, con el polinomio $(x,y)\to x^2+y^2$). Es inmediato ver que

$$Df(x,y)=(0,0)\iff x^2+y^2=\frac{1}{\frac{\pi}{4}+k\pi}, k\in\mathbb{N}\cup\{0\},\ \text{ o }\ (x,y)=(0,0).$$

Así pues los conjuntos $C_k = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1/(\pi/4 + k\pi)\}$ están formados todos ellos por puntos críticos de f, y f es constante sobre cada C_k (donde vale $(-1)^k e^{-(\pi/4+k\pi)}\sqrt{2}/2$, $k = 0, 1, 2, ...$). El origen es también un punto crítico de f, y no hay más. Es fácil ver que f tiene máximos relativos sobre cada punto de C_k con k par, y mínimos relativos sobre cada punto de C_k con k impar, mientras que el origen es un punto de silla. Por otro lado, f es constante en ∂K (donde vale $e^{-1/5}\sin(1/5)$). De todo esto se concluye, comparando los valores

$$0,\ e^{-1/5}\sin(1/5),\ (-1)^k e^{-(\pi/4+k\pi)}\sqrt{2}/2\ (k \in \mathbb{N} \cup \{0\}),$$

que f alcanza su máximo absoluto sobre K en cada punto de C_0, y su mínimo en cada punto de C_1. De hecho, puesto que

$$\lim_{(x,y)\to\infty} f(x,y) = 0,$$

estos mínimos y máximos de f en K lo son también de f en todo $\mathbb{R}^2$.

Los ejemplos anteriores eran muy simples en cuanto al comportamiento de f en la frontera de K, porque f era constante en ∂K, y así su mínimo y máximo en la frontera eran iguales y se alcanzaban en cualquier punto de ésta. Supongamos ahora que no hemos tenido tanta suerte y que f no es constante en ∂K. ¿Qué podemos hacer para determinar los números

$$\alpha = \inf\{f(x) : x \in \partial U\} \ \text{ y } \ \beta = \sup\{f(x) : x \in \partial U\}\ ?$$

Una primera observación que puede hacerse es que a veces no es necesario calcular α y β, sino que puede bastar con acotarlos, si conseguimos una cota que nos permita concluir, cotejándola con otros valores de f en el interior de K, que el máximo y el mínimo de f en K no pueden alcanzarse en la frontera ∂K y deben por tanto hallarse en $U = \text{int}(K)$.

Ejemplo 4.3. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función del ejemplo anterior, y $K = \{(x,y) \in \mathbb{R}^2 : x^2 + 3y^2 \leq 25\}$. Ahora f ya no es constante en $\partial K = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 25\}$, pero, teniendo en cuenta que la función $t \mapsto e^{-1/t}\sin(1/t)$ es creciente para $t > 4/\pi$, podemos acotar α y β como sigue. Para todo $(x,y) \in \partial K$ tenemos que $x^2 + y^2 \leq x^2 + 3y^2 = 25$, y por tanto

$$e^{-1/(x^2+y^2)}\sin\frac{1}{x^2+y^2} \leq e^{-1/(x^2+3y^2)}\sin\frac{1}{x^2+3y^2} = e^{-1/25}\sin\frac{1}{25},$$

de donde se deduce que

$$\beta := \sup\{f(x,y) : (x,y) \in \partial K\} \leq e^{-1/25}\sin\frac{1}{25}.$$

Análogamente, teniendo en cuenta que si $(x,y) \in \partial K$ entonces $x^2 + y^2 \geq \frac{1}{3}(x^2 + 3y^2) = 25/3 > 8$, se deduce que

$$\alpha := \inf\{f(x,y) : (x,y) \in \partial K\} \geq e^{-1/8}\sin\frac{1}{8}.$$

Por tanto, puesto que

$$\beta \leq e^{-1/25}\sin\frac{1}{25} < f(C_0) = e^{-\pi/4}\frac{\sqrt{2}}{2},$$

y que

$$f(C_1) = -e^{-\pi/4-\pi}\sin\frac{\sqrt{2}}{2} < 0 < e^{-1/8}\sin\frac{1}{8} \leq \alpha,$$

es obvio que f alcanza su máximo absoluto sobre K en cada uno de los puntos de C_0, y su mínimo absoluto sobre K en los puntos de C_1.

Supongamos ahora que no somos capaces de acotar α y β satisfactoriamente para ver que el mínimo y el máximo deben alcanzarse en el interior de K, o incluso que tenemos la sospecha de que el mínimo o el máximo de hecho deben alcanzarse en la frontera ∂K. ¿Cómo podemos intentar calcular α y β, y hallar los puntos donde se alcanzan?

Ejemplo 4.4. Sean $f(x,y) = x + y + 1$, $K = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Calcular el máximo y el mínimo de f sobre K.

La función f no tiene ningún punto crítico en $\mathbb{R}^2$. Por tanto su máximo y mínimo sobre K no pueden alcanzarse en el interior de K, sino que deben estar forzosamente en la frontera. Puesto que la frontera de K es una curva diferenciable, la circunferencia $\partial K = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, podemos parametrizar esta curva y estudiar los mínimos y máximos de la composición de dicha parametrización con f, que deberán corresponder con los de f en ∂K. En efecto, sea $\gamma : [0, 2\pi] \to \partial K$ definida por

$$\gamma(t) = (\cos t, \sin t).$$

La curva γ es diferenciable y sobreyectiva. Por tanto

$$\max_{t\in[0,2\pi]} f(\gamma(t)) = \max_{(x,y)\in\partial K} f(x,y),$$

luego, analizando los puntos críticos de $g(t) = f(\gamma(t))$, podemos determinar el máximo de g en $[0, 2\pi]$ y por tanto también el de f en ∂K, y lo mismo puede decirse de los mínimos. Tenemos que

$$g'(t) = -\sin t + \cos t = 0 \iff t \in \{\pi/4, 3\pi/4\},$$

y como $g(\pi/4) = f(\sqrt{2}/2, \sqrt{2}/2) = 1 + \sqrt{2}$ y $g(3\pi/4) = f(-\sqrt{2}/2, -\sqrt{2}/2) = 1 - \sqrt{2}$, es claro que el valor máximo de f en ∂K (y por tanto en K) es $1 + \sqrt{2}$ y se alcanza en $(\sqrt{2}/2, \sqrt{2}/2)$, y el mínimo es $1 - \sqrt{2}$ y se alcanza en $(\sqrt{2}/2, \sqrt{2}/2)$.

Ejemplo 4.5. Hallar los extremos absolutos de $f(x,y) = x - x^2 - y^2$, en $K = [-1,1] \times [-1,1]$.

En este caso f sólo tiene un punto crítico en $(1/2, 0)$, que corresponde a un máximo local. Por tanto, al menos el mínimo de f en K debe alcanzarse en la frontera. Para hallar los extremos de f en ∂K podemos parametrizar los cuatro segmentos que componen ∂K y componer f con estas curvas, obteniendo así cuatro funciones,

$$h_1(t) = f(1,t);\; h_2(t) = f(-1,t);\; h_3(t) = f(t,1)\;,\; h_4(t) = f(t,-1),$$

$t \in [-1,1]$, de las cuales podemos calcular sus extremos en el intervalo $[-1,1]$. El problema se reduce entonces a comparar los valores de $f \circ h_i$ en los puntos críticos de las funciones h_i con los valores que toma f en el punto $(1/2, 0)$ y en los vértices del cuadrado K. El resultado es que el máximo de f en K es $1/4$ y se alcanza en el punto $(1/2, 0)$, mientras que el mínimo es -3 y se alcanza en los vértices $(-1, 1)$ y $(-1, -1)$.

Cuando se use esta técnica de parametrizar las curvas que componen el borde de un compacto K, es importante asegurar que las parametrizaciones γ sean regulares, es decir, que $\gamma'(t) \neq 0$ para todo t. De lo contrario, entre los puntos críticos de la función $g(t) = f(\gamma(t))$ estarían todos los $\gamma(t)$ para los cuales $\gamma'(t) = 0$ (ya que $g'(t) = Df(\gamma(t))(\gamma'(t))$, por la regla de la cadena), que no tienen por qué ser extremos de la restricción de f a γ.

En dimensiones superiores los bordes de los dominios que aparecen con frecuencia en problemas de este tipo van a ser trozos de superficies (o hipersuperficies) diferenciables. En lugar de parametrizarlos como en el caso de curvas, resulta más conveniente utilizar el siguiente resultado (que es el caso particular más importante de un teorema más general conocido como de los multiplicadores de Lagrange).

Teorema 4.1. *Sean $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ funciones de clase C^1. Sean $x_0 \in U$, $r = g(x_0)$, $S_r = g^{-1}(r)$ el conjunto de nivel r de g. Supongamos que $\nabla g(x_0) \neq 0$, y que la restricción de f a S_r, denotada $f_{|S_r}$, tiene un máximo o un mínimo en x_0. Entonces existe $\lambda \in \mathbb{R}$ tal que*

$$\nabla f(x_0) = \lambda \nabla g(x_0).$$

Es decir, los extremos de $f_{|S_r}$ se alcanzan en puntos en los que el gradiente de f es perpendicular a la superficie de nivel S_r de g.

Demostración. Por un lado, recuérdese que el espacio tangente a S_r en x_0 se define como el espacio ortogonal a $\nabla g(x_0)$, y la justificación de esto fue precisamente el hecho de que $\nabla g(x_0)$ es perpendicular al vector tangente $\gamma'(0)$ a cualquier camino diferenciable $\gamma : (-1, 1) \to S_r$ tal que $\gamma(0) = x_0$; además, para todo vector v del hiperplano tangente a S_r en x_0 (es decir, para todo vector v perpendicular a $\nabla g(x_0)$), puesto que por el teorema de la función implícita S_r puede expresarse localmente como la gráfica de una función que depende diferenciablemente de $n - 1$ coordenadas de $\mathbb{R}^n$, existe ciertamente una curva diferenciable $\gamma : (-1, 1) \to S_r$ tal que $\gamma'(0) = v$. Por otro lado, si γ es una de estas curvas, la función $h(t) = f(\gamma(t))$ debe tener un extremo en $t = 0$, y por tanto

$$0 = h'(0) = \langle \nabla f(x_0), \gamma'(0) \rangle,$$

es decir, $\nabla f(x_0)$ es perpendicular a todos los vectores $\gamma'(0)$. Así $\nabla f(x_0)$ es también ortogonal al espacio tangente a S_r en x_0. Por tanto $\nabla f(x_0)$ y $\nabla g(x_0)$ deben ser paralelos (al ser ambos perpendiculares al mismo hiperplano), es decir, existe $\lambda \in \mathbb{R}$ tal que $\nabla f(x_0) = \lambda \nabla g(x_0)$. □

Ejemplo 4.6. Hallar los extremos de $f(x, y, z) = ax + by + cz$ en $K = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}$.

Salvo en el caso trivial en que $a = b = c = 0$, f no tiene ningún punto crítico en el interior de K (ni en $\mathbb{R}^3$), luego los extremos de f en K deben alcanzarse en ∂K. Definamos $g(x, y, z) = x^2 + y^2 + z^2$. Nótese que $\partial K = g^{-1}(1)$ y que $\nabla g(x, y, z) \neq 0$ para todo $(x, y, z) \in \partial K$. Si (x, y, z) es uno de los extremos, deberá satisfacer, según el teorema anterior, que, para algún $\lambda \in \mathbb{R}$,

$$(a, b, c) = \nabla f(x, y, z) = \lambda \nabla g(x, y, z) = \lambda(2x, 2y, 2z),$$

es decir (x, y, z) es proporcional a (a, b, c). Como además es $x^2 + y^2 + z^2 = 1$, se deduce que

$$(x, y, z) \in \left\{ \frac{1}{\sqrt{a^2 + b^2 + c^2}}(a, b, c), \ -\frac{1}{\sqrt{a^2 + b^2 + c^2}}(a, b, c) \right\}.$$

Es obvio entonces que el máximo de f en ∂K es

$$\frac{a + b + c}{\sqrt{a^2 + b^2 + c^2}},$$

y se alcanza en el punto $\frac{1}{\sqrt{a^2+b^2+c^2}}(a, b, c)$, mientras que el mínimo es

$$-\frac{a + b + c}{\sqrt{a^2 + b^2 + c^2}},$$

y se alcanza en $-\frac{1}{\sqrt{a^2+b^2+c^2}}(a, b, c)$.

La versión general del teorema de los multiplicadores de Lagrange es el siguiente resultado.

Teorema 4.2 (de los multiplicadores de Lagrange)**.** *Sean $F = (F_1, ..., F_k) : U \subseteq \mathbb{R}^n \to \mathbb{R}^k$ una función de clase C^1, y supongamos que rango$(DF(x)) = k$ para todo $x \in \mathbb{R}^n$ con $F(x) = 0$. Sea $f : U \to \mathbb{R}$ diferenciable en U. Si a es un punto de máximo o mínimo de $f(x)$ sujeta a las restricciones $F_1(x) = ... = F_k(x) = 0$, entonces*

$$\nabla f(a) = \lambda_1 \nabla F_1(a) + ... + \lambda_k \nabla F_k(a)$$

para ciertos números $\lambda_1, ..., \lambda_k \in \mathbb{R}$.

Demostración. Sea $S = \{x \in U : F(x) = 0\}$, que por el teorema de la función implícita puede expresarse localmente como gráfica de una función φ que depende diferenciablemente de $n - k$

coordenadas de $\mathbb{R}^n$ y toma valores en las restantes coordenadas: por ejemplo, puede suponerse, salvo reordenamiento de la base canónica de $\mathbb{R}^n = \mathbb{R}^{n-k} \times \mathbb{R}^k$, que en un entorno W de a se tiene $S \cap W = \{x = (u, v) \in (\mathbb{R}^{n-k} \times \mathbb{R}^k) \cap W : v = \varphi(u)\}$. Y también podemos expresar S como

$$S = \bigcap_{j=1}^{k} S_j,$$

donde

$$S_j = \{x \in U : F_j(x) = 0\}.$$

Para cada curva diferenciable $\gamma : (-1, 1) \to S$ con $\gamma(0) = a$, puesto que $\gamma(t) \in S_j$ para cada j y $\nabla F_j(a)$ es perpendicular a S_j en a, se tiene que $\gamma'(0)$ y $\nabla F_j(a)$ son perpendiculares. Por tanto, $\gamma'(0)$ es perpendicular al subespacio de dimensión k generado por $\nabla F_1(a), ..., \nabla F_k(a)$, que denotaremos N_a. Esto prueba que el conjunto de todos los vectores que son de la forma $v = \gamma'(0)$ para alguna de estas curvas γ, y que denotaremos T_a, es ortogonal a N_a. Por otra parte, para cada una de estas curvas, como f tiene un extremo en a, se cumple que

$$0 = \frac{d}{dt} f(\gamma(t))_{|t=0} = \langle \nabla f(a), \gamma'(0) \rangle,$$

lo que muestra que $\nabla f(a)$ también es ortogonal a T_a. Ahora bien, es fácil ver que T_a coincide con el subespacio vectorial de dimensión $n - k$ definido por $T_a = \{(u, D\varphi(a)(u)) : u \in \mathbb{R}^{n-k}\}$. Entonces necesariamente tenemos la descomposición $\mathbb{R}^n = T_a \oplus N_a$ con factores ortogonales, y por tanto $\nabla f(a) \in T_a^{\perp} = N_a$, lo que significa que $\nabla f(a)$ es combinación lineal de $\nabla F_1(a), ..., \nabla F_k(a)$. $\square$

Por supuesto puede ocurrir que la frontera de un dominio K de $\mathbb{R}^3$ en el cual deseamos maximizar o minimizar una función no sea una sola hipersuperficie, sino por ejemplo una unión finita de superficies que están *pegadas* unas a otras a lo largo de curvas (que a su vez podrían estar pegadas unas a otras mediante puntos que tampoco pueden ignorarse). Entonces habría que usar, en cada en cada una de estas curvas, y en cada una de estas superficies, los métodos descritos anteriormente, para posteriormente comparar los valores de la función en los extremos condicionados así hallados con los valores de la función en los puntos de *pegado* y también con los valores críticos de la función correspondientes a puntos críticos del interior de K.

Más en general, en $\mathbb{R}^n$, con $n \geq 4$, podríamos tener una unión finita de hipersuperficies $(n-1)$-dimensionales que están pegadas unas a otras a lo largo de uniones de superficies $(n-2)$-dimensionales a su vez pegadas... . Como se ve la complejidad del problema aumenta mucho con la dimensión.

También, por supuesto, puede ocurrir que la frontera del conjunto K no sea descriptible como unión finita de trozos de hipersuperficies diferenciables, en cuyo caso no podemos dar ningún método general para calcular extremos.

Problemas

Problema 4.1. Estudiar los puntos críticos de las siguientes funciones, tratando de determinar si corresponden a máximos locales, mínimos locales, o puntos de silla:

1. $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = e^{xy}$;

2. $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x, y, z) = x^2 + y^2 + z^2 + 4yz + y^2x - z^3$;

3. $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x, y, z) = x^3 + y^3 + z^3 - 3(xy + xz + yz)$;

4. $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^3 + y^3 - 3ax$, según los valores de $a \in \mathbb{R}$;

5. $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2$;

6. $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2 - y^4$.

Problema 4.2. Estúdiense los puntos críticos, máximos y mínimos relativos de las siguientes funciones:

$$\begin{array}{rl}
f(x,y) = x - x^2 - y^2 & g(x,y) = x^2 - y^2 + xy \\
h(x,y) = x^2 + y^2 + 3xy & F(x,y) = y^2 - x^3 \\
H(x,y) = x^3 + y^3 - 3axy & \psi(x,y) = \sin(x^2 + y^2) \\
\varphi(x,y) = xye^{x+2y} & \Phi(x,y) = e^{1+x^2-y^2} \\
\phi(x,y,z) = x^2 + y^2 - z^2 & P(x,y,z) = x^2 + y^2 + z^2 - 2x + yz
\end{array}$$

Problema 4.3. Calcúlense:

1. Los extremos absolutos de las funciones siguientes, en los conjuntos que se indican:

   a) $f(x,y) = x - x^2 - y^2$, en $K = \{(x,y) \in \mathbb{R}^2, \quad \text{máx}\{\mid x \mid, \mid y \mid\} \leq 1\}$.

   b) $f(x,y) = \sin(xy)$, en $K = \{(x,y) \in \mathbb{R}^2, \quad \text{máx}\{\mid x \mid, \mid y \mid\} \leq 1\}$.

   c) $f(x,y) = \sin(x+y) + \cos(x-y)$, en $K = [0, 2\pi] \times [0, 2\pi]$.

2. $\sup\{4x^2 + y^2 - 4x - 3y;\ y \geq 0,\ 4x^2 + y^2 \leq 4\}$.

3. $\sup\{(x+y)e^{-(x^2+y^2)};\ (x,y) \in \mathbb{R}^2\}$.

Problema 4.4. Sea $f(x,y) = (y - 3x^2)(y - x^2)$

1. Pruébese que (0,0) es un punto crítico de f.

2. Pruébese que f tiene un mínimo relativo en (0,0) sobre cada recta que pasa por (0,0).

3. Pruébese que (0,0) no es un mínimo relativo de f, sino un punto de silla.

Problema 4.5. Encuéntrense los puntos $(x,y) \in \mathbb{R}^2$ y las direcciones para las que la derivada direccional de $f(x,y) = 3x^2 + y$ alcanza el máximo valor, suponiendo que (x,y) verifica $x^2 + y^2 \leq 1$.

Problema 4.6. Sea $f(x,y) = a[2xy + y^2 + yx^2 + \cos(x+y)] + x^2(a^2 - y)$. Discútase la existencia de extremos en el origen, según los valores de a.

Problema 4.7. Encuéntrense los extremos relativos y los absolutos, si existen, de las siguientes funciones:

1. $f(x,y) = (ax^2 + by^2)e^{-(x^2+y^2)}, \quad a, b > 0$.

2. $f(x,y,z) = (x + z^2)e^{x(y^2+z^2+1)}$.

Problema 4.8. Hállense los extremos de $f(x,y,z) = xy^2z^3(a - x - y - z)$.

Problema 4.9. Hallar los puntos de la superficie $z^2 = xy + 1$ de $\mathbb{R}^3$ que están más próximos al origen $(0,0,0)$.

Problema 4.10. Probar que el máximo de $x^2y^2z^2$ sobre la esfera $x^2 + y^2 + z^2 = r^2$ es $(r^2/3)^3$.

Problema 4.11. Probar que el mínimo de $x^2 + y^2 + z^2$ sobre la superficie $x^2y^2z^2 = (r^2/3)^3$ es r^2.

Problema 4.12. Encontrar el máximo de $x_1^2x_2^2...x_n^2$ sobre la esfera $x_1^2 + x_2^2 + ... + x_n^2 = 1$. Deducir que para cualesquiera n números positivos a_i se tiene que

$$(a_1a_2...a_2)^{1/n} \leq \frac{a_1 + a_2 + ... + a_n}{n},$$

es decir, la media geométrica es siempre menor o igual que la media aritmética.

Problema 4.13. Sean $a_1, ..., a_n$ números positivos. Para $p > 1$, sea q tal que

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Hallar el máximo de $f(x) = \sum_{i=1}^{n} a_i x_i$ sobre la superficie S definida por $\sum_{i=1}^{n} |x_i|^q = 1$. Deducir que

$$\sum_{i=1}^{n} a_i x_i \leq \left(\sum_{i=1}^{n} a_i^p \right)^{1/p} \left(\sum_{i=1}^{n} |x_i|^q \right)^{1/q}$$

(la *desigualdad de Hölder* para sumas).

Problema 4.14. Sea $f : \mathbb{R} \to \mathbb{R}$ diferenciable en $\mathbb{R}$ tal que f tiene un mínimo local en un punto x_0, y que no hay más puntos críticos. Probar que x_0 es entonces un mínimo global.

Problema 4.15. Probar que el problema anterior no es válido para funciones $f : \mathbb{R}^2 \to \mathbb{R}$. Considérese para ello la función $f(x, y) = x^2 + y^2(1 - x)^3$, mostrando que $(0, 0)$ es el único punto crítico, que es un mínimo local estricto, y que f no tiene ningún mínimo global.