Diferenciabilidad de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

Cálculo II (2004) *

En este capítulo generalizamos la noción de diferenciabilidad para funciones vectoriales de variable vectorial, que también llamamos *aplicaciones*. Una aplicación es entonces una función $f\colon X \to \mathbb{R}^m$, donde X es un subconjunto de $\mathbb{R}^n$.

Ejemplo 1 *(Transformaciones Lineales).* Un ejemplo central de aplicaciones son las *transformaciones lineales*, es decir, las aplicaciones $T\colon \mathbb{R}^n \to \mathbb{R}^m$ que verifican $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ para α, β reales, x, y vectores en $\mathbb{R}^n$, todos arbitrarios. Recordemos que dada una transformación lineal T como antes existe una única matriz $A = ((a_{ij}))$ de tamaño $m \times n$, que llamamos *matriz asociada*, tal que se verifica

$$T(x) = Ax^t,$$

asumiendo que $x = (x_1, \dots, x_n)$ es un vector fila. Si las funciones coordenadas de T son $T_1, \dots, T_m$, tenemos $T_i(x) = \langle A^i, x\rangle$, para cada $i = 1, \dots, m$ donde $A^i = (a_{i1}, \dots, a_{in})$ designa la i-ésima fila de la matriz A.

Otros casos importantes de aplicaciones son los *cambios de coordenadas*, como los que vemos a continuación.

Ejemplo 2 *(Coordenadas Polares).*
El cambio de coordenadas cartesianas (x, y) a *coordenadas polares* (ρ, θ), viene dado por la aplicación

$$x = \rho\cos\theta, \qquad y = \rho\operatorname{sen}\theta, \tag{1}$$

que definimos del dominio $X = [0, \infty) \times [0, 2\pi)$ en $\mathbb{R}^2$.

*Notas para el curso de la Licenciatura en Matemática, Facultad de Ciencias, preparadas por Ernesto Mordecki.

Ejemplo 3 *(Coordenadas cilíndricas).* Si en el ejemplo anterior agregamos la "altura" z de un punto (x, y, z) de $\mathbb{R}^3$, obtenemos

$$x = \rho\cos\theta, \quad y = \rho \operatorname{sen}\theta, \quad z = z, \tag{2}$$

aplicación definida de $X = [0, \infty) \times [0, 2\pi) \times \mathbb{R}$ en $\mathbb{R}^3$, que llamamos cambio de *coordenadas cilíndricas*.

Ejemplo 4 *(Coordenadas esféricas).* Ahora describimos un punto (x, y, z) de $\mathbb{R}^3$ mediante sus *coordenadas esféricas*, designadas (ρ, θ, φ), de forma que

$$x = \rho\cos\theta \operatorname{sen}\varphi, \quad y = \rho \operatorname{sen}\theta \operatorname{sen}\varphi, \quad z = \rho\cos\varphi, \tag{3}$$

aplicación definida de $X = [0, \infty) \times [0, 2\pi) \times [0, \pi)$ en $\mathbb{R}^3$.

1. Aplicaciones diferenciables

Definición 1 (Aplicación diferenciable). *Sean $f : U \to \mathbb{R}^m$, U abierto en $\mathbb{R}^n$. La aplicación f es* diferenciable *en un punto $a \in U$, cuando existen una transformación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$ y una aplicación $p : B^*(0, \varepsilon) \to \mathbb{R}^m$, con $\varepsilon > 0$ suficientemente pequeño, tales que para todo $v \in \mathbb{R}^n$ con $a + v \in B^*(a, \varepsilon)$, se verifica*

$$f(a + v) - f(a) = T(v) + \|v\|\, p(v), \quad \lim_{v\to 0} p(v) = 0. \tag{4}$$

La transformación lineal anterior se llama diferencial *de la aplicación f en el punto a, y se designa df_a. Decimos además que f es* diferenciable en U *cuando es diferenciable en todo $a \in U$.*

Introduciendo la aplicación $r(v) = \|v\|p(v)$, que llamamos *resto*, las fórmulas en (4) pueden escribirse como

$$f(a + v) - f(a) = T(v) + r(v), \qquad \lim_{v\to 0} \frac{r(v)}{\|v\|} = 0.$$

Estamos entonces definiendo que una aplicación es diferenciable en un punto a cuando su incremento se puede aproximar por una transformación lineal. Observemos además que esta definición generaliza la de diferenciabilidad de funciones de varias variables (caso $m = 1$) del capítulo anterior.

Si ponemos $v = 0$ en (4) vemos que p se puede definir en $v = 0$ poniendo $p(0) = 0$. Consideremos entonces, de aquí en más, $p\colon B(0,\varepsilon) \to \mathbb{R}^m$.

Como las transformaciones lineales son continuas, es inmediato obtener que si f es diferenciable en un punto a entonces es continua: tenemos $f(a+v) \to f(a)$ si $v \to 0$, como se ve tomando límite a la derecha en (4).

Ejemplo 5. Veamos que una transformación lineal $T\colon \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable. Tenemos

$$T(a+v) - T(a) = T(v),$$

por lo que se verifica la definición (4), con $dT_a = T$, y $p(v) = 0$ para todo $v \in \mathbb{R}^n$.

Como una aplicación $f\colon U \to \mathbb{R}^m$ equivale a sus m funciones coordenadas $f_1, \dots, f_m$, es natural relacionar la diferenciabilidad de una aplicación con la de sus funciones coordenadas. Las coordenadadas son funciones reales de variable vectorial, cuya diferenciabilidad definimos en el capítulo anterior.

Teorema 1. *Consideremos $f = (f_1, \dots, f_m)\colon U \to \mathbb{R}^m$, $a \in U$, abierto de $\mathbb{R}^n$. Entonces la aplicación f es diferenciable en a si y sólo si sus coordenadas $f_1, \dots, f_m$ son diferenciables en a.*

Demostración. Supongamos primero que la aplicación f es diferenciable en a, y escribamos la identidad vectorial (4) como m identidades escalares. Designando $T = (T_1, \dots, T_m)$, $p = (p_1, \dots, p_m)$, tenemos

$$f_i(a+v) - f_i(a) = T_i(v) + \|v\|\, p_i(v), \qquad \lim_{v \to 0} p_i(v) = 0,$$

para $i = 1, \dots, m$, que es la definición de diferenciabilidad de la función f_i en a, dado que $T_i(v) = \langle A^i, v\rangle$, donde A^i es la i-ésima fila de la matriz asociada a T.

Reciprocamente, si $f_1, \dots, f_m$ son diferenciables en a, existen funciones $p_1, \dots, p_m$ tales que

$$f_i(a+v) - f_i(a) = df_{ia}(v) + \|v\|\, p_i(v), \qquad \lim_{v \to 0} p_i(v) = 0,$$

para $i = 1, \dots, m$. Luego, con

$$T(v) = \big(df_{1a}(v), \dots, df_{ma}(v)\big), \tag{5}$$

y $p = (p_1, \dots, p_m)$, la aplicación f verifica la definición (4) en el punto a, concluyendo la demostración. □

Ejemplo 6 (Curva diferenciable). Una curva $\lambda = (\lambda_1, \dots, \lambda_n) \colon I \to \mathbb{R}^n$ es diferenciable en un punto a del intervalo real I si son diferenciables sus coordenadas. Obtenemos entonces la misma definición de *curva diferenciable* del capítulo anterior. Además, si λ es diferenciable en $a \in I$, tenemos

$$\frac{\lambda(a+h) - \lambda(a)}{h} = d\lambda_a(1) + \frac{|h|}{h} p(h),$$

con $p(h) \to 0$ $(h \to 0)$, de donde resulta, que

$$\lambda'(a) = d\lambda_a(1), \tag{6}$$

es decir, la velocidad de la curva en el instante a se obtiene como el valor del diferencial de la curva en a, evaluado en el vector $v = 1$.

La matriz asociada al diferencial df_a de una aplicación f en un punto a se llama *matriz jacobiana*, o más brevemente *jacobiano*, y se designa mediante $Jf(a)$. De la fórmula (5) obtenemos, si $f = (f_1, \dots, f_m)$ es diferenciable en a, que se verifica

$$df_a(v) = \big(df_{1a}(v), \dots, df_{ma}(v)\big), \tag{7}$$

y como $df_{ia}(v) = \langle \nabla f_i(a), v \rangle$, resulta que la fila i-ésima de $Jf(a)$ es el gradiente $\nabla f_i(a)$. En otros términos, hemos demostrado que

$$Jf(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix}. \tag{8}$$

De aquí resulta además la unicidad del diferencial de una aplicación diferenciable en un punto.

Definimos la *derivada direccional* de una aplicación f en un punto a con respecto de un vector $v \in \mathbb{R}^n$, que designamos $(\partial f/\partial v)(a)$, como el valor del límite en $\mathbb{R}^m$ dado por

$$\frac{\partial f}{\partial v}(a) = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t},$$

cuando existe. A partir de la definición resulta que la derivada direccional de una aplicación existe, si y sólo si existen las derivadas direccionales de sus funciones coordenadas. Además, en caso de existencia, se verifica

$$\frac{\partial f}{\partial v}(a) = \Big(\frac{\partial f_1}{\partial v}(a), \dots, \frac{\partial f_m}{\partial v}(a)\Big). \tag{9}$$

Pero $(\partial f_i/\partial v)(a) = df_{ia}(v)$, por lo que el vector obtenido en (9), en vista de (7), es la imagen de v por el diferencial de f, es decir

$$\frac{\partial f}{\partial v}(a) = df_a(v).$$

La fórmula anterior es una generalización a aplicaciones de la que obtuvimos para funciones reales de varias variables.

Teorema 2 (Regla de la Cadena para aplicaciones).
Consideremos las aplicaciones $f\colon U \to \mathbb{R}^m$ donde $U \subset \mathbb{R}^n$, $g\colon V \to \mathbb{R}^p$, donde $f(U) \subset V \subset \mathbb{R}^m$, con U, V abiertos. Supongamos que f es diferenciable en $a \in U$, y que g es diferenciable en $b = f(a)$. Entonces, la función compuesta $h = g \circ f\colon U \to \mathbb{R}^p$ es diferenciable en el punto a. Mas aún, el diferencial de h en a es la composición del diferencial de g en b con el diferencial de f en a, es decir

$$dh_a = dg_b \circ df_a, \tag{10}$$

lo que expresado a través de las matrices jacobianas da la fórmula

$$Jh(a) = Jg(b) \times Jf(a), \tag{11}$$

donde $\times$ designa el producto de matrices.

Antes de la demostración veamos que este teorema generaliza la regla de la cadena de funciones del capítulo anterior. Supongamos para esto que $p = 1$. Según vimos, $f = (f_1, \dots, f_m)$ es diferenciable si y sólo si $f_1, \dots, f_m$ son diferenciables, y las hipótesis de ambos teoremas coinciden si $p = 1$. En este caso el jacobiano de g es una matriz $1 \times m$, y viene dado por

$$Jg(b) = \Big[\frac{\partial g}{\partial y_1}(b) \cdots \frac{\partial g}{\partial y_m}(b)\Big]. \tag{12}$$

Por su parte, el jacobiano de h es una matriz $1 \times n$, dada por

$$Jh(a) = \Big[\frac{\partial h}{\partial x_1}(a) \cdots \frac{\partial h}{\partial x_n}(a)\Big]. \tag{13}$$

Es claro entonces que multiplicando $Jg(b)$ dado en (12) por $Jf(a)$ dado en (8), que es igual a $Jh(a)$ dado en (13), obtenemos las n fórmulas para las derivadas parciales de la función compuesta de la regla de la cadena para funciones.

Demostración. La demostración es análoga a la de la regla de la cadena para funciones, pero (paradójicamente) mas sencilla, facilitada por el cálculo matricial.

Como f es diferenciable en a, existe $p\colon B(0,\varepsilon)\to\mathbb{R}^m$ tal que, si $a+v\in B(0,\varepsilon)$ tenemos

$$f(a+v)-f(a)=df_a(v)+\|v\|\,p(v),\qquad \lim_{v\to 0}p(v)=0. \tag{14}$$

Como g es diferenciable en $b=f(a)$, existe $q\colon B(0,\delta)\to\mathbb{R}^p$ tal que

$$g(b+w)-g(b)=dg_b(w)+\|w\|\,q(w),\qquad \lim_{v\to 0}q(v)=0.$$

Sustituyendo $w=f(a+v)-f(a)$ dado en (14) en la fórmula anterior, tenemos

$$h(a+v)-h(a)=g(b+w)-g(b)=dg_b\big(df_a(v)+\|v\|\,p(v)\big)+\|w\|\,q(w).$$

Introduciendo la función auxiliar $P(v)$, podemos entonces escribir

$$h(a+v)-h(a)=\big(dg_b\circ df_a\big)(v)+\|v\|\,P(v), \tag{15}$$

$$P(v)=dg_b\big(p(v)\big)+\frac{\|w\|}{\|v\|}q(w). \tag{16}$$

Si demostramos que $P(v)\to 0$ $(v\to 0)$, en vista de (15) resulta que h es diferenciable en a, y que su diferencial verifica (10). En primer lugar, como la norma de $v/\|v\|$ es uno, tenemos

$$\frac{\|w\|}{\|v\|}\leq\left\|df_a\Big(\frac{v}{\|v\|}\Big)\right\|+\|p(v)\|\leq|||df_a|||+\|p(v)\|\,,$$

que es una función acotada si $v\to 0$. Luego $w\to 0$ cuando $v\to 0$ y el segundo sumando a la derecha en (16) tiende a cero. El primero tiende a cero dado que dg_b es continua, y $p(v)\to 0$. Esto concluye la demostración. □

Observación. Una demostración alternativa de la regla de la cadena para aplicaciones se basa en la regla de la cadena para funciones. Demostramos que las coordenadas de h son diferenciables, y obtenemos la fórmula (11) a partir de las fórmulas para el cálculo de las derivadas.

La regla de la cadena recién demostrada nos permite obtener una *interpretación geométrica* del diferencial de una aplicación. Sea $f\colon U\to\mathbb{R}^m$ diferenciable en $a\in U$, abierto de $\mathbb{R}^n$. Dado un vector $v\in\mathbb{R}^n$, considerando la curva $\lambda(t) = a + tv$, en el dominio, obtenemos la curva en el recorrido $\mu(t) = f\big(\lambda(t)\big)$. La velocidad de esta curva en el punto $\mu(0) = b = f(a)$, está dada por

$$\mu'(0) = d\mu_0(1) = d\big(f \circ \lambda\big)_0(1) = df_a \circ d\lambda_0(1) = df_a(v),$$

donde aplicamos (6). Si suponemos entonces que un punto se encuentra en $\lambda(t)$ en el instante t, mientras la función f nos da la imagen de su posición, el diferencial df_a nos da la velocidad a la que se mueve esta imagen.

2. Teorema de la aplicación inversa

Comencemos con algunas definiciones relativas a la regularidad de aplicaciones y de sus inversas. Decimos que una aplicación es de clase C^k cuando sus funciones coordenadas son de clase C^k para $k = 0, 1, \ldots, \infty$.

Definición 2. (a) *Una aplicación* $f\colon X \to Y$*, donde* $X \subset \mathbb{R}^n$*,* $Y \subset \mathbb{R}^m$*, es un* homeomorfismo *cuando es biyectiva, continua en* X*, y su aplicación inversa* $f^{-1}\colon Y \to X$ *es continua en* Y*. Cuando existe un homeomorfismo* $f\colon X \to Y$ *decimos que* X *e* Y *son* homeomorfos*.*
(b) *Una aplicación* $f\colon U \to V$*, donde* $U \subset \mathbb{R}^n$*,* $V \subset \mathbb{R}^m$*, ambos abiertos, es un* difeomofismo *cuando es biyectiva, diferenciable en* U*, y su aplicación inversa* $f^{-1}\colon V \to U$ *es diferenciable en* V*. Cuando existe un difeomorfismo* $f\colon U \to V$ *decimos que* U *y* V *son* difeomorfos*.*
(c) *Una aplicación* $f\colon X \to \mathbb{R}^m$*, donde* $X \subset \mathbb{R}^n$*, es un* difeomorfismo local *en* $a \in X$ *cuando existen dos conjuntos abiertos* U, V*, tales que* $a \in U \subset X$*,* $V \subset \mathbb{R}^m$*, de forma que la restricción de* f *a* U *en* V*, es decir,* $f\colon U \to V$*, es un difeomorfismo.*
(d) *Un* C^k*-difeomorfismo es un difeomorfismo tal que él y su inverso son de clase* C^k*, para* $k = 1, \ldots, \infty$*.*

Observemos que la composición de homeomorfismos es un homeomorfismo (porque la composición de aplicaciones continuas es continua), y que la composición de difeomorfismos es un difeomorfismo (por la regla de la cadena).

Es también sencillo de verificar que si $h = g \circ f$, donde f es un difeomorfimso local en un punto a, y g es un difeomorfismo local en el punto $b = f(a)$, entonces h es un difeomorfismo local en a.

Además, un difeomorfismo es siempre un homeomorfismo, y un homeomorfismo es siempre biyectivo, siendo falsos ambos recíprocos. En particular, la función $f(x) = x^3$ es un homeomorfismo, es diferenciable, pero no es un difeomorfismo.

Con estas definiciones, nuestro problema es obtener condiciones para que una aplicación de clase C^1 sea un difeomorfismo, o, por lo menos, un difeomorfismo local.

Recordemos que dada una función de una variable $f\colon I \to \mathbb{R}$, donde I es un intervalo de $\mathbb{R}$, si f es derivable y $f'(x) > 0$ en I, obtenemos que existe su función inversa $f^{-1}\colon f(I) \to I$, que es derivable, y su derivada vale

$$\big(f^{-1}\big)'(y) = \frac{1}{f'\big(f^{-1}(y)\big)},$$

como resulta del teorema de la función inversa en $\mathbb{R}$.

En este capítulo los resultados son mas modestos, dado que, en general, no es posible invertir una aplicación en todo su dominio. Como el rol de la derivada lo juega el diferencial, que es una transformación lineal, comenzamos recordando cuando una transformación lineal es invertible.

Ejemplo 7. Sea $T\colon \mathbb{R}^n \to \mathbb{R}^m$ una transformación lineal. Recordemos que si $m \neq n$ la transformación T no puede ser invertible. Si $m = n$ la transformación lineal T es invertible si y sólo si $\det(T) \neq 0$. Además, en este caso la aplicación inversa es también una transformación lineal, su matriz asociada es la matriz inversa de la matriz de T, por lo que es diferenciable. Es decir, $T\colon \mathbb{R}^n \to \mathbb{R}^n$ con $\det(T) \neq 0$ es un difeomorfismo.

Veamos ahora dos propiedades de la norma $|||T|||$ de una transformación lineal que serán útiles en lo que sigue. En primer lugar, si $v \in \mathbb{R}^n$ no es nulo, poniendo $x = v/\left\|v\right\|$ tenemos

$$\left\|T(v)\right\| = \left\|v\right\| \times \left\|T(x)\right\| \leq |||T||| \times \left\|v\right\|. \tag{17}$$

De aquí resulta que las transformaciones lineales son funciones lipchitzianas, con constante $|||T|||$.

Veamos ahora que, si $A = ((a_{ij}))$ es la matriz de T, tenemos

$$|||T||| \leq mn \max\{|a_{ij}|\colon i = 1, \dots, m; j = 1, \dots, n\}. \tag{18}$$

En efecto, si $x = x_1e_1 + \cdots + x_ne_n$, y $\left\|x\right\| = 1$, tenemos $|x_i| \leq 1$ $(i = 1, \dots, n)$, por lo que

$$\begin{aligned}\|T(x)\| &\leq |x_1|\,\|T(e_1)\| + \cdots + |x_n|\,\|T(e_n)\| \leq \|T(e_1)\| + \cdots + \|T(e_n)\| \\ &\leq m\big(\|T(e_1)\|_\infty + \cdots + \|T(e_n)\|_\infty\big) \\ &\leq mn \max\{|a_{ij}|\colon i = 1, \ldots, m; j = 1, \ldots, n\}.\end{aligned}$$

El teorema del valor medio no es válido en el contexto de las aplicaciones. Sin embargo, vale el siguiente resultado.

Teorema 3 (Desigualdad del valor medio).
Sea $f\colon U \to \mathbb{R}^m$ con U abierto en $\mathbb{R}^n$. Consideremos un punto $a \in U$ y un vector v de $\mathbb{R}^n$, tal que el intervalo $[a, a+v]$ esté contenido en U.
(a) *Supongamos que f restringida a $[a, a+v]$ es una función continua, y que existe la derivada direccional $(\partial f/\partial v)(x)$ para todo $x \in (a, a+v)$. Entonces,*

$$\|f(a+v) - f(a)\| \leq \sup_{0<\theta<1} \left\| \frac{\partial f}{\partial v}(a+\theta v) \right\|. \tag{19}$$

(b) *Cuando f es diferenciable para cada $x \in [a, a+v]$, se verifica*

$$\|f(a+v) - f(a)\| \leq \Big(\sup_{0<\theta<1} |||df_{a+\theta v}||| \Big) \|v\|. \tag{20}$$

Demostración. Dado $w \in \mathbb{R}^m$, la función auxiliar $\phi\colon [0,1] \to \mathbb{R}$ definida mediante $\phi(t) = \langle w, f(a+tv)\rangle$ es continua, por serlo f restringida a $[a, a+v]$. Tenemos además

$$\frac{\phi(\theta+h) - \phi(\theta)}{h} = \left\langle w, \frac{f(a+\theta v + hv) - f(a+\theta v)}{h} \right\rangle \to \left\langle w, \frac{\partial f}{\partial v}(a+\theta v) \right\rangle,$$

si $h \to 0$, por lo que ϕ es derivable en $(0,1)$. Aplicando entonces el teorema de Lagrange obtenemos que existe $\theta \in (0,1)$ tal que $\phi(1) - \phi(0) = \phi'(\theta)$. Como $\phi(1) = \langle w, f(a+v)\rangle$, $\phi(0) = \langle w, f(a)\rangle$, tomando $w = f(a+v) - f(a)$ y aplicando la desigualdad de Cauchy-Schwarz, resulta

$$\begin{aligned}\|f(a+v) - f(a)\|^2 &= \langle w, f(a+v) - f(a)\rangle = \phi(1) - \phi(0) = \phi'(\theta) \\ &= \left\langle w, \frac{\partial f}{\partial v}(a+\theta v) \right\rangle \leq \|w\| \left\| \frac{\partial f}{\partial v}(a+\theta v) \right\| \\ &\leq \|f(a+v) - f(a)\| \sup_{0<\theta<1} \left\| \frac{\partial f}{\partial v}(a+\theta v) \right\|,\end{aligned}$$

de donde obtenemos (19).

Veamos ahora (b). Como f es diferenciable en los puntos $x \in [a, a+v]$ estamos en las hipótesis de la parte (a). Además, como $(\partial f/\partial v)(x) = df_x(v)$, aplicando la desigualdad (17) obtenemos

$$\|df_x(v)\| \leq |||df_x||| \, \|v\| \,,$$

por lo que (20) se obtiene de (19), concluyendo la demostración. □

El siguiente ejemplo muestra que no es posible obtener igualdad en el teorema del valor medio para aplicaciones.

Ejemplo 8. Consideremos la curva $\lambda \colon \mathbb{R} \to \mathbb{R}^2$ dada por $\lambda(t) = (t^2, t^3)$, y el intervalo $[0,1] = [a, a+v]$ con $a = 0$, $v = 1$. Si existe $\theta \in (0,1)$ tal que

$$(1,1) = \lambda(1) - \lambda(0) = \frac{\partial \lambda}{\partial v}(\theta) = \lambda'(\theta) = (2\theta, 3\theta^2),$$

necesariamente se verifica, simúltaneamente, $\theta = 1/2$, y $\theta = 1/\sqrt{3}$, lo que es imposible. Este ejemplo muestra que no es posible obtener un teorema del valor medio con igualdad en el contexto de las aplicaciones.

Estudiemos ahora algunas condiciones necesarias que verifican las aplicaciones cuando son difeomorfismos.

Teorema 4. *Consideremos un difeomorfismo $f \colon U \to V$, con inversa dada por $f^{-1} \colon V \to U$, donde $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ son abiertos. Entonces, se verifica $m = n$ y, para cada $a \in U$ con $b = f(a)$, los diferenciales df_a y $d(f^{-1})_b$ son transformaciones lineales invertibles e inversas, es decir*

$$d(f^{-1})_b = \big(df_a\big)^{-1}, \tag{21}$$

lo que en términos de matrices jacobianas se traduce en

$$J(f^{-1})(b) = \big[Jf(a)\big]^{-1}.$$

Demostración. Como el diferencial de la identidad $\mathrm{Id}_U \colon U \to U$ es la transformación lineal identidad $\mathrm{Id}_n \colon \mathbb{R}^n \to \mathbb{R}^n$, aplicando la regla de la cadena en la igualdad $\mathrm{Id}_U = f^{-1} \circ f$, obtenemos

$$\mathrm{Id}_n = d\big(\mathrm{Id}_U\big)_a = d\big(f^{-1} \circ f\big)_a = d(f^{-1})_b \circ df_a.$$

En conclusión, la transformación lineal $d(f^{-1})_b$ es invertible, y tiene inversa df_a, probando (21). Además, como df_a es una transformación lineal, resulta $m = n$, concluyendo la demostración. □

El siguiente teorema muestra que un homeomorfismo f diferenciable en un punto a, con diferencial df_a invertible, tiene aplicación inversa diferenciable en $b = f(a)$.

Teorema 5 (Diferenciabilidad de la inversa de un homeomorfismo). *Sea $f\colon U \to V$ un homeomorfismo entre los conjuntos abiertos $U \subset \mathbb{R}^n$ y $V \subset \mathbb{R}^m$.*
(a) *Si f es diferenciable en $a \in U$ y el diferencial df_a es una transformación lineal invertible, entonces $f^{-1}\colon V \to U$ es diferenciable en $b = f(a)$.*
(b) *Si f es diferenciable en U y df_x es invertible para todo $x \in U$, entonces f es un difeomorfismo.*

Demostración. Observemos primero que (b) es consecuencia directa de (a). Veamos entonces la demostración de (a).

Según vimos, si f^{-1} es diferenciable en b, su diferencial es $\big(df_a\big)^{-1}$. Escribimos entonces la condición de diferenciabilidad para f^{-1} en b, definiendo

$$q(w) = \frac{f^{-1}(b+w) - f^{-1}(b) - \big(df_a\big)^{-1}(w)}{\|w\|}. \tag{22}$$

Se trata entonces de verificar que $q(w) \to 0$ $(w \to 0)$. Tenemos $a = f^{-1}(b)$. Designando $v = f^{-1}(b+w) - f^{-1}(b)$, tenemos $a + v = f^{-1}(b+w)$, y como f es diferenciable en a, podemos escribir

$$w = b + w - b = f(a+v) - f(a) = df_a(v) + \|v\|\, p(v), \qquad \lim_{v\to 0} p(v) = 0.$$

Sustituyendo esta última expresión en (22), tenemos

$$\begin{aligned} q(w) &= \frac{f^{-1}(b+w) - f^{-1}(b) - \big(df_a\big)^{-1}\big(df_a(v) + \|v\|\, p(v)\big)}{\|w\|} \\ &= -\big(df_a\big)^{-1}\Big(\frac{\|v\|}{\|w\|}\Big)p(v). \end{aligned} \tag{23}$$

Como f y f^{-1} son continuas obtenemos que $v \to 0$ si y sólo si $w \to 0$. Entonces, si $w \to 0$ resulta que $p(v) \to 0$, y por ser $\big(df_a\big)^{-1}$ continua, es suficiente ver que el cociente $\|v\| \,/\, \|w\|$ permanece acotado si $w \to 0$. Veamos que $\inf\{\|df_a(x)\| : \|x\| = 1\} = m > 0$.En efecto, de anularse este ínfimo, aplicando el teorema de Weierstrass (la función df_a es continua, y el dominio $\{\|x\| = 1\}$ es compacto) resulta que existe $x \neq 0$ tal que $df_a(x) = 0$, lo que contradice que df_a es invertible. Entonces, como $p(v) \to 0$ si $w \to 0$, obtenemos, en vista de (23), que

$$\frac{\|w\|}{\|v\|} = \left\| df_a\Big(\frac{v}{\|v\|}\Big) + p(v) \right\| \geq \left\| df_a\Big(\frac{v}{\|v\|}\Big) \right\| - \|p(v)\| \geq \frac{m}{2},$$

en un entorno adecuado del origen. Tenemos entonces que $\|v\| / \|w\| \leq m/2$, obteniendo que $q(w) \to 0$ si $w \to 0$, concluyendo la demostración. $\square$

Para demostrar el teorema de la función inversa precisamos el siguiente resultado, basado en la completitud de $\mathbb{R}^n$, que fundamenta el *método de las aproximaciones sucesivas*. Dados $X \subset \mathbb{R}^n$ y $g\colon X \to X$ definimos la sucesión $\big(g^{(k)}(x)\big)$ de los *iterados* de $x \in X$ por g de la siguiente forma:

$$g^{(0)}(x) = x, \qquad g^{(k)}(x) = g\big(g^{(k-1)}(x)\big) \quad (k = 1, 2, \dots).$$

Teorema 6 (de la contracción).
Consideremos un subconjunto X cerrado de $\mathbb{R}^n$, y una aplicación $g\colon X \to X$. Supongamos que g es una contracción*, es decir, existe $\lambda \in (0,1)$ tal que*

$$\|g(x) - g(y)\| \leq \lambda \|x - y\|, \qquad \textit{para todos } x, y \in X.$$

Entonces, existe un único $z \in X$ tal que $g(z) = z$, que llamamos punto fijo *de g. Además*

$$z = \lim_k g^{(k)}(x), \tag{24}$$

para todo $x \in X$.

Demostración. Veamos primero que si existe un punto fijo, este es único. En efecto, si $z_1 = g(z_1)$, $z_2 = g(z_2)$, tenemos

$$\|z_1 - z_2\| = \|g(z_1) - g(z_2)\| \leq \lambda \|z_1 - z_2\|,$$

de donde $\|z_1 - z_2\| = 0$, y resulta $z_1 = z_2$.

Para verificar la existencia de un punto fijo, tomemos cualquier $x \in X$. Si $x \neq g(x)$ veriquemos que la sucesión de los iterados de x es una sucesión de Cauchy en X. Sea entonces $\varepsilon > 0$. Para dos índices $i < j = i + p$, tenemos

$$\begin{aligned} \big\|g^{(j)}(x) - g^{(i)}(x)\big\| &= \big\|g^{(i)}\big(g^{(p)}(x)\big) - g^{(i)}(x)\big\| \\ &\leq \lambda^i \big\|g^{(p)}(x) - x\big\|. \end{aligned}$$

Por otra parte

$$\begin{aligned}\left\|x - g^{(p)}(x)\right\| &\leq \|x - g(x)\| + \left\|g(x) - g^{(2)}(x)\right\| + \cdots + \left\|g^{(p-1)}(x) - g^{(p)}(x)\right\| \\ &\leq \left(1 + \lambda + \cdots + \lambda^{p-1}\right) \|x - g(x)\| \leq \frac{1}{1-\lambda} \|x - g(x)\| .\end{aligned}$$

En conclusión

$$\left\|g^{(j)}(x) - g^{(i)}(x)\right\| \leq \frac{\lambda^i}{1-\lambda} \|x - g(x)\| < \varepsilon,$$

si $i \geq i_0$ con i_0 suficientemente grande. Obtenemos entonces que la sucesión $\left(g^{(k)}(x)\right)$ verifica la condición de Cauchy, por lo que existe $z = \lim g^{(k)}(x)$, y $z \in X$ por ser X cerrado. Finalmente, basados en la continuidad de g (que es lipchitziana por ser una contracción), tenemos

$$z = \lim_k g^{(k)}(x) = \lim_k g\left(g^{(k-1)}(x)\right) = g(z),$$

verificando que z es el punto fijo que buscábamos. □

Estamos ahora en condiciones de demostrar el teorema más importante del capítulo.

Teorema 7 (Teorema de la aplicación inversa).
Consideremos $f\colon U \to \mathbb{R}^n$ de clase C^1 en U, abierto de $\mathbb{R}^n$. Si el diferencial de f en a es una transformación lineal invertible, es decir, si $\det\left(Jf(a)\right) \neq 0$, *entonces f es un difeomorfismo local en a, es decir, existen dos abiertos U_0, V_0 que verifican $a \in U_0 \subset U$, $V_0 \subset \mathbb{R}^n$, y tales que $f\colon U_0 \to V_0$ es un difeomorfismo.*

Demostración. Comencemos con la demostración en el caso particular en el que

$$a = 0, \quad f(a) = 0, \quad df_a = \mathrm{Id}_n, \tag{25}$$

para luego obtener el caso general a partir de éste.

Consideremos para ésto la aplicación auxiliar

$$g(x) = x - f(x),$$

definida en U. Esta aplicación tiene diferencial nulo en el origen porque $dg_0 = \mathrm{Id}_n - df_0 = 0$, por lo que la matriz jacobiana es nula, y todas las derivadas parciales de las coordenadas de g son nulas, y por ser continuas, tienden a cero si $x \to 0$. Esto implica, en vista de (18), que $|||dg_x||| \to 0$ $(x \to 0)$, por lo que existe una bola cerrada $B[0, 2\varepsilon] \subset U$ tal que

$$|||dg_x||| \leq \frac{1}{2} \quad \text{para todo } x \in B[0, 2\varepsilon]. \tag{26}$$

Aplicando la desigualdad del valor medio obtenemos, si $x, y \in B[0, 2\varepsilon]$, que se verifica

$$\|g(x) - g(y)\| \leq \sup_{0<\theta<1} |||dg_{x+\theta y}||| \, \|x - y\| \leq \frac{1}{2} \|x - y\|, \tag{27}$$

lo que, en particular, si $y = 0$, nos da

$$\|g(x)\| \leq \frac{1}{2} \|x\|.$$

Obtenemos entonces que $g\colon B[0, 2\varepsilon] \to B[0, \varepsilon]$, y es una contracción con constante $1/2$.

Definimos la aplicación inversa en la bola abierta $V_0 = B(0, \varepsilon)$. Tomemos $y \in V_0$ para ver que existe un único $x \in B[0, 2\varepsilon]$ tal que $y = f(x)$.

Para esto trasladamos según y la aplicación g, definiendo, para $x \in B[0, 2\varepsilon]$ la aplicación

$$g_y(x) = y + g(x) = y + x - f(x).$$

Se verifica

$$\|g_y(x)\| \leq \|y\| + \|g(x)\| < \varepsilon + \frac{1}{2} \|x\| \leq 2\varepsilon$$

por lo que $g_y\colon B[0, 2\varepsilon] \to B[0, 2\varepsilon]$. La acotación (27) nos permite ver que g_y también es una contracción. En efecto

$$\|g_y(x_1) - g_y(x_2)\| \leq \|g(x_1) - g(x_2)\| \leq \frac{1}{2} \|x_1 - x_2\|. \tag{28}$$

En conclusión existe un único $x \in B[0, 2\varepsilon]$ tal que $g_y(x) = x$, es decir, un único $x \in B[0, \varepsilon]$ tal que $y = f(x)$, como queríamos verificar. Además, el conjunto

$$U_0 = f^{-1}(V_0) \subset B[0, 2\varepsilon] \subset U$$

es abierto, por ser f continua y V_0 abierto (Teorema del capítulo 1). Estas consideraciones nos permiten concluir que $f\colon U_0 \to V_0$ es una función biyectiva. Nos resta ver que $f^{-1}\colon V_0 \to U_0$ es continua y diferenciable.

Sean entonces $y_1, y_2 \in V_0$, y $x_1 = f^{-1}(y_1)$, $x_2 = f^{-1}(y_2)$. Dado que $x = g(x) + f(x)$ tenemos

$$\begin{aligned}\|x_2 - x_1\| &\leq \|g(x_2) - g(x_1)\| + \|f(x_2) - f(x_1)\| \\ &\leq \frac{1}{2}\|x_2 - x_1\| + \|f(x_2) - f(x_1)\|,\end{aligned}$$

obteniendo que

$$\left\|f^{-1}(y_2) - f^{-1}(y_1)\right\| = \|x_2 - x_1\| \leq 2\|f(x_2) - f(x_1)\| = 2\|y_2 - y_1\|.$$

de donde f^{-1} resulta Lipchitziana, por lo tanto continua.

Para concluir que f es un difeomorfismo, observemos que $\det\big(Jf(0)\big) = 1$, y por ser f de clase C^1 existe una bola, que suponemos $B[0, 2\varepsilon]$, en donde $\det\big(Jf(x)\big) \neq 0$. Aplicamos entonces (b) en el teorema 5 para obtener que f es un difeomorfismo en U_0. Esto concluye la demostración en el caso particular (25).

Veamos por último la demostración en el caso general. Dada f en las condiciones del teorema, consideramos $g(z) = z + a$ y

$$\tilde{f}(z) = \big(df_a\big)^{-1}\big[f\big(g(z)\big) - f(a)\big].$$

La aplicación $\tilde{f}$ verifica las condiciones del teorema, y además verifica (25), por lo que es un difeomorfismo local en $z = 0$. Además

$$f(x) = \big(df_a\big) \circ \tilde{f}\big(g^{-1}(x)\big) + f(a).$$

Es decir, f es un difeomorfismo local en a por ser composición de difeomorfismos locales. Esto conlcuye la demostración $\square$

3. Teorema de la aplición implícita y multiplicadores de Lagrange

En esta sección vemos primero como se aplica el teorema de la aplicación inversa para "despejar" variables $y = (y_1, \dots, y_m)$ en función de otras $x = (x_1, \dots, x_n)$, cuando existe una relación entre ellas dada por un sistema de ecuaciones $F_i(x_1, \dots, x_n, y_1, \dots, y_m) = 0$ $(i = 1, \dots, m)$, relación que en notación vectorial escribimos $F(x, y) = 0$, si $F = (F_1, \dots, F_m)$.

Para enunciar en nuestro caso la condición equivalente a $F_y(x_0, y_0) \neq 0$ del teorema de la función implícita en dos variable, introducimos la notación

$$\frac{\partial F}{\partial y} = \frac{\partial(F_1, \ldots, F_m)}{\partial(y_1, \ldots, y_m)} = \begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{bmatrix},$$

que es el jacobiano de la aplicación de la variable $y = (y_1, \ldots, y_m)$, que se obtiene cuando $x = (x_1, \ldots, x_n)$ permanece constante. Designamos además

$$(x_0, y_0) = (x_{10}, \ldots, x_{n0}, y_{10}, \ldots, y_{m0}).$$

Teorema 8 (Teorema de la aplicación implícita).
Sean $(x_0, y_0) \in U$, abierto en $\mathbb{R}^{n+m}$, $F = (F_1, \ldots, F_m) \colon U \to \mathbb{R}^m$ de clase C^1, tales que $F(x_0, y_0) = 0$. Supongamos que $\det\big(\partial F/\partial y\big)(x_0, y_0) \neq 0$. *Entonces:*
(a) *Existen bolas $B_n = B(x_0, \delta)$, $B_m = B(y_0, \varepsilon)$ tales que $B_n \times B_m \subset U$, y una función $f \colon B_n \to B_m$ de clase C^1 tal que*

$$F(x, y) = 0 \Leftrightarrow y = f(x) \text{ para todo } (x, y) \in B_n \times B_m.$$

(b) *Si $x \in B_n$ el jacobiano de f verifica*

$$Jf(x) = -\Big[\frac{\partial F}{\partial y}\big(x, f(x)\big)\Big]^{-1} \frac{\partial F}{\partial x}\big(x, f(x)\big). \tag{29}$$

Demostración. Para utilizar el teorema de la aplicación inversa, consideremos $\varphi \colon U \to \mathbb{R}^{n+m}$ definido mediante

$$\varphi(x, y) = \big(x, F(x, y)\big). \tag{30}$$

La aplicación φ es diferenciable en U por serlo sus coordenadas y, derivando, obtenemos que su jacobiano, escrito en bloques, vale

$$J\varphi = \begin{bmatrix} \mathrm{Id}_n & 0 \\ \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \end{bmatrix}.$$

Para calcular el determinante de este jacobiano desarrollamos por las primeras n filas, y obtenemos

$$\det(J\varphi)(x_0, y_0) = \det\Big(\frac{\partial F}{\partial y}\Big)(x_0, y_0) \neq 0,$$

según nuestra hipótesis, por lo que del teorema de la aplicación inversa obtenemos que existe una bola $B_{n+m} = B\big((x_0, y_0), \varepsilon\big)$ tal que φ restringida a B_{n+m} es un difeomorfismo, con aplicación inversa que designamos $\psi(u,v) = \big(\psi_1(u,v), \dots, \psi_n(u,v), \dots, \psi_{n+m}(u,v)\big)$. Observemos que, según la forma que tiene φ en (30), donde sus primeras coordenadas son la identidad, esta propiedad se obtiene también para ψ, que por lo tanto se puede escribir como

$$\psi(u,v) = \big(u, h(u,v)\big)$$

donde $h = (\psi_{n+1}, \dots, \psi_{n+m})$. Veamos ahora que $f(x) = h(x,0)$ verifica (a), para $x \in B_n = B(x_0, \delta)$, con $\delta > 0$ suficientemente pequeño, tal que $\psi\big(B\big((x_0,0),\delta\big)\big) \subset B_{n+m}$. En efecto, como $(x,0) \in B\big((x_0,0),\delta\big)$ porque $x \in B(x_0,\delta)$, y tenemos

$$\begin{aligned}\big(x, F\big(x, f(x)\big)\big) &= \varphi(x, f(x)) = \varphi\big(x, h(x,0)\big) \\ &= \varphi\big(\psi(x,0)\big) = (x,0),\end{aligned}$$

obteniendo que

$$F\big(x, f(x)\big) = 0. \tag{31}$$

Veamos ahora (b). La aplicación f es diferenciable, por estar formada por coordenadas de una aplicación diferenciable, mientras que F es de clase C^1. Designando $g(x) = \big(x, f(x)\big)$, tenemos $F \circ g = 0$, y podemos aplicar la regla de la cadena en (31), para obtener $JF \circ Jg = 0$. Si escribimos los jacobianos en bloques tenemos

$$JF = \frac{\partial(F_1,\dots,F_m)}{\partial(x_1,\dots,x_n,y_1,\dots,y_m)} = \begin{bmatrix} \frac{\partial(F_1,\dots,F_m)}{\partial(x_1,\dots,x_n)} & \frac{\partial(F_1,\dots,F_m)}{\partial(y_1,\dots,y_m)} \end{bmatrix},$$

$$Jg = \frac{\partial(x_1,\dots,x_n,f_1,\dots,f_m)}{\partial(x_1,\dots,x_n)} = \begin{bmatrix} \mathrm{Id}_n \\ \frac{\partial(f_1,\dots,f_m)}{\partial(x_1,\dots,x_n)} \end{bmatrix}.$$

Imponiendo la condición $JF \times Jg = 0$ obtenemos

$$\frac{\partial(F_1,\dots,F_m)}{\partial(x_1,\dots,x_n)} \times \mathrm{Id}_n + \frac{\partial(F_1,\dots,F_m)}{\partial(y_1,\dots,y_m)} \times \frac{\partial(f_1,\dots,f_m)}{\partial(x_1,\dots,x_n)} = 0,$$

de donde resulta (b), por ser $\frac{\partial(F_1,\dots,F_m)}{\partial(y_1,\dots,y_m)}$ una matriz invertible. Finalmente observamos que las derivadas de f se obtienen a partir de las derivadas de F, que son continuas, evaluadas en x y en $f(x)$. Como f es continua, por ser diferenciable, y la composición de aplicaciones continuas es continua, resulta que f es de clase C^1. Esto concluye la demostración. □

Ejemplo 9. Queremos estudiar el conjunto M de puntos $(x, y, z) \in \mathbb{R}^3$ que verifican

$$\begin{cases} x^2 + y^2 = 1 \\ y = z \end{cases}$$

Tenemos $(1, 0, 0) \in M$. Consideremos la aplicación $F = (F_1, F_2)\colon \mathbb{R}^3 \to \mathbb{R}^2$ dada por

$$F(x, y, z) = (x^2 + y^2 - 1, y - z).$$

Observamos que $m+n = 3$ y $m = 2$, es decir, podemos despejar dos variables, digamos x, y en función de la tercera z. Como $F \in C^1$, resta verificar la condición de no anulación del determinante del jacobiano

$$\frac{\partial(F_1, F_2)}{\partial(x, y)} = \begin{bmatrix} 2x & 2y \\ 0 & 1 \end{bmatrix}$$

En el punto $(1, 0, 0)$ tenemos $det\big(\partial(F_1, F_2)/\partial(x, y)\big) = 2 \neq 0$, por lo que aplicando el teorma de la aplicación implícita obtenemos que existe una función $f\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^2$, dada por $f(z) = \big(x(z), y(z)\big)$ tal que se verifica $F\big(x(z), y(z), z\big) = 0$. En otros términos M es (localmente) la curva $\lambda\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ de ecuación $\lambda(z) = \big(x(z), y(z), z\big)$. Calculemos el vector tangente a esta curva en $z = 0$. Tenemos $\partial F/\partial z = \big[(F_1)_z, (F_2)_z\big]^t$. Por eso, en el punto $(1, 0, 0)$, tenemos

$$\big(x'(0), y'(0)\big) = \frac{\partial(x, y)}{\partial z}(1, 0, 0) = -\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \times \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

De aquí concluímos que $\lambda'(0) = (0, 1, 1)$, que es el vector tangente a la curva M en el punto $(1, 0, 0)$.

Teorema 9 (Multiplicadores de Lagrange).
Consideremos U abierto en $\mathbb{R}^{n+m}$, $f\colon U \to \mathbb{R}$, $g = (g_1, \dots, g_m)\colon U \to \mathbb{R}^m$, ambas funciones de clase C^1.
(a) *Si $p \in M = g^{-1}(0)$ y f presenta extremo relativo condicionado a g en p, entonces el conjunto de vectores*

$$\{\nabla f(p), \nabla g_1(p), \dots, \nabla g_m(p)\} \tag{32}$$

es linealmente dependiente.
(b) *Si además $Jg(p)$ tiene rango m (rango máximo), entonces existen reales $\lambda_1, \dots, \lambda_m$, llamados* multiplicadores de Lagrange*, tales que*

$$\nabla f(p) = \lambda_1 \nabla g_1(p) + \cdots + \lambda_m \nabla g_m(p).$$

Demostración. Comencemos observando que (b) es inmediato a partir de (a). Supongamos entonces, para demostrar (a), por absurdo, que el conjunto de $m+1$ vectores en (32) es linealmente independiente. Obtenemos entonces que la matriz

$$\frac{\partial(f, g_1, \dots, g_m)}{\partial(x_1, \dots, x_{n+m})} = \begin{bmatrix} \nabla f \\ \nabla g_1 \\ \vdots \\ \nabla g_m \end{bmatrix}$$

tiene sus $m+1$ filas linealmente independientes, y por lo tanto, tiene rango $m+1$. Tiene entonces también $m+1$ columnas linealmente independientes, que por simplicidad en la notación, suponemos son las $m+1$ primeras. Designemos $p = (p_1, \dots, p_m, p_{m+1}, \dots, p_{n+m})$. Fijando entonces las variables $p_{m+2}, \dots, p_{m+n}$ obtenemos que la aplicación auxiliar $y = (y_0, \dots, y_m) = \varphi(x_1, \dots, x_{m+1})$ dada por

$$\begin{aligned} y_0 &= f(x_1, \dots, x_{m+1}, p_{m+2}, \dots, p_{m+n}) \\ y_1 &= g_1(x_1, \dots, x_{m+1}, p_{m+2}, \dots, p_{m+n}) \\ &\vdots \\ y_m &= g_m(x_1, \dots, x_{m+1}, p_{m+2}, \dots, p_{m+n}) \end{aligned}$$

tiene jacobiano no nulo en el punto $a = (p_1, \dots, p_{m+1})$, y por lo tanto, existe una bola $B(a, \varepsilon)$ en donde φ es un difeomorfismo. Ademas $\varphi(a) = \big(f(p), g_1(p), \dots, g_m(p)\big) = \big(f(p), 0, \dots, 0\big)$. Como existe la inversa de φ, para $\delta > 0$ suficientemente pequeño, los puntos $\big(f(p) \pm \delta, 0, \dots, 0\big)$ tienen preimagen en la bola $B(a, \varepsilon)$, y por este motivo, estos puntos pertenecen a $g^{-1}(0)$, y f no presenta extremo relativo condicionado en p, contradiciendo nuestra hipótesis. En conclusión, la matriz no tiene rango $m+1$, y los vectores en (32) son linealmente dependientes, como queríamos probar. □

Al igual que en la búsquda de los extermos de f con una condición $g = 0$, el teorema de los multiplicadores de Lagrange da una condición necesaria para la existencia de extremos relativos con varias condiciones $g = (g_1, \dots, g_m) = 0$. Una vez descartados los puntos de $M = g^{-1}(0)$ en los que f, g no son de clase C^1, y aquellos en los que Jg no tiene rango máximo, los extremos condicionados se presentan únicamente en aquellos puntos en los que existen $\lambda_1, \dots, \lambda_m$ reales tal que

$$\nabla f(p) = \lambda_1 \nabla g_1(p) + \cdots + \lambda_m \nabla g_m(p).$$

Para buscar estos puntos consideramos la función auxiliar

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f - \lambda_1 g_1 - \cdots - \lambda_m g_m,$$

y observamos que los extremos condicionados de f a g son puntos críticos de la función L. Veamos una aplicación de este procedimiento.

Ejemplo 10. Queremos determinar los puntos del conjunto M del ejemplo 9 cuya distancia al origen es máxima y mínima. Para eso consideramos la función $f(x, y, z) = x^2 + y^2 + z^2$ (cuyos extremos coinciden con los de la distancia al origen) y la condición $g(x, y, z) = (x^2 + y^2 - 1, y - z) = 0$. Tenemos entonces

$$L(x, y, z, \lambda, \mu) = x^2 + y^2 + z^2 - \lambda(x^2 + y^2 - 1) - \mu(y - z).$$

Las ecuaciones que se obtienen de imponer $\nabla L = 0$ son

$$\begin{cases} L_x = 2x(1-\lambda) = 0 \\ L_y = 2y(1-\lambda) - \mu = 0 \\ L_z = 2z + \mu = 0 \end{cases} \qquad \begin{cases} L_\lambda = -x^2 - y^2 + 1 = 0 \\ L_\mu = -y + z = 0 \end{cases}$$

Si sustituímos $y = z$ y observamos que $x = 0$ ó $\lambda = 1$, obtenemos todas las soluciones de este sistema, que son $(1, 0, 0)$, $(-1, 0, 0)$, $(0, 1, 1)$ y $(0, -1, -1)$. Evualuando la funcón en estos puntos obtenemos que la distancia mínima es 1, que se da en los puntos $(\pm 1, 0, 0)$, y la máxima $\sqrt{2}$ que se presenta en los puntos $(0, \pm 1, \pm 1)$.