UNIVERSIDAD NACIONAL DE INGENIERÍA

Facultad de Ciencias

Escuela Profesional de Matemática

Tesis para optar el Título Profesional de:
Licenciado en Matemática

Títulada:
Desigualdades Variacionales, Soluciones y Aplicaciones

Presentada por:
Bach. Miguel Angel Paz Arcelles

Asesor:
Dr. Pedro Canales García

LIMA-PERÚ
2011

Índice general

Desigualdades Variacionales, Soluciones y Aplicaciones

Resumen

En este trabajo presentamos la teoría sobre Desigualdades Variacionales (DV's). En el Capítulo 1 damos una breve reseña de la Teoría sobre DV's y los casos que se presentan. En el Capítulo 2 hacemos referencia a ejemplos que ilustran diversas formulaciones de una DV. En el Capítulo 3 se resumen los resultados más importantes de optimización no diferenciable debido a que en los problemas de aplicación con frecuencia podemos encontrar funciones con un número finito de puntos de no diferenciabilidad. En el Capítulo 4 se presentan los métodos posibles de solución a una DV. En el Capítulo 5 se dan las conclusiones.

Capítulo 1

Preliminares

1.1. El Problema de Desigualdad Variacional

Sea $X \subseteq \mathbb{R}^n$ un conjunto convexo, cerrado y no vacío; $u : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función convexa, propia[1] y semi-continua inferiormente[2] (lsc - abreviatura en inglés); y $F : \operatorname{dom} u \cap X \to \mathbb{R}^n$ un mapeo continuo de valor vectorial en $\operatorname{dom} u \cap X$[3].

Para definir nuestro problema, vamos a hacer uso de tres operadores:
El Cono Normal de X en el punto x

$$N_X(x) = \begin{cases} \{z \in \mathbb{R}^n \mid z^T(y - x) \leq 0, \ \ \forall\, y \in X\}, & x \in X \\ \emptyset, & x \notin X; \end{cases} \tag{1.1}$$

El Subdiferencial para u

$$\partial u(x) = \{\xi_u \in \mathbb{R}^n \mid u(y) \geq u(x) + \xi_u^T(y - x), \ \ \forall y \in \mathbb{R}^n\};$$

y el mapeo F. Consideremos entonces el problema de hallar una vector $x^* \in \mathbb{R}^n$ tal que

$$0^n \in F(x^*) + \partial u(x^*) + N_X(x^*) \tag{1.2}$$

Este problema es conocido como una *Desigualdad Variacional Generalizada* y lo vamos a denotar como GVIP(F, u, X), y a su conjunto de soluciones como SOL(F, u, X).

[1] La función u es propia si $u(x) < +\infty$ para al menos algún x y $u(x) > -\infty$ para todo x

[2] La función u es lsc en x_0 si los valores de u en puntos cercanos a x_0 son cercanos a $u(x_0)$ o mayores que $u(x_0)$

[3] El dominio efectivo $\operatorname{dom} u$ de u es el subconjunto de $\mathbb{R}^n$ para el cual $u(x) < +\infty$

1.2. Instancias del Problema

Bajo las siguientes suposiciones, el problema GVIP(F, u, X) puede ser equivalentemente definido usando la función u en lugar de su mapeo subdifencial.

Condición 1.1 (De Regularidad). int(dom $u) \cap X \neq \emptyset$ ■

Observación 1.1. Sabiendo que la *Función Indicadora* δ_X para un conjunto X es

$$\delta_X = \begin{cases} 0, & x \in X, \\ +\infty, & x \notin X. \end{cases} \tag{1.3}$$

entonces, según las definiciones que hemos dado, el cono normal N_X asociado con el conjunto convexo X es el subdiferencial de la función indicadora δ_X para X.
Usando la condición 1.1 garantizamos que $\partial[u + \delta_X](x) = \partial u(x) + \partial \delta_X(x)$, con $x \in \text{dom}\, u \cap X$.
Si el dom$u = \mathbf{R}^n$, se ve satisfecha la condición 1.1 . ■

Proposición 1.2 (Equivalencia de Desigualdad Variacional).
Bajo la condición 1.1, el problema GVIP(F, u, X) es equivalente al problema de buscar un $x^* \in X$ tal que

$$F(x^*)^T(x - x^*) + u(x) - u(x^*) \geq 0, \quad \forall x \in X \tag{1.4}$$

Prueba .- Consideremos el problema convexo

$$\underset{x \in X}{\text{minimizar}}\ h(x) = F(x^*)^T x + u(x) \tag{1.5}$$

donde $x^* \in X$. Es claro que (1.4) es equivalente a que x^* sea la solución óptima global al problema en (1.5). Tomando en cuenta la condición 1.1, podemos caracterizar a x^* como

$$0^n \in \partial h(x^*) + N_X(x^*) \tag{1.6}$$

Por definición de $h(x)$ y de subdiferencial, tenemos que $\partial[F(x^*)^T x + u(x)] \subset F(x^*) + \partial u(x)$, y por la condición 1.1 obtenemos la otra inclusión. Así,

$$\partial h(x) = F(x^*) + \partial u(x), \qquad x \in X; \tag{1.7}$$

y combinando (1.6) y (1.7) obtenemos el resultado deseado. ■

Corolario 1.3 (Formulación Equivalente de Desigualdad Variacional). Bajo la condición 1.1, el problema GVIP(F, u, X) es equivalente al problema de hallar un $x^* \in X$ tal que

$$F(x^*)^T(x - x^*) + u'(x^*; x - x^*) \geq 0, \qquad \forall\, x \in X \tag{1.8}$$

<u>*Prueba*</u>.- El resultado se sigue de considerar que $h'(x^*; x - x^*) \geq 0$ para todo $x \in X$ es una condición necesaria y suficiente para que x^* sea solución óptima de (1.5). ∎

Ejemplo 1.2 (Sistema de Desigualdades Variacionales sobre un producto cartesiano). Sea el conjunto factible de GVIP(F, u, X) descrito por un *Producto Cartesiano* de conjuntos factibles,

$$X = \prod_{i \in C} X_i, \qquad X_i \subseteq \mathbf{R}^{n_i}, \qquad \sum_{i \in C} n_i = n, \tag{1.9}$$

para algún conjunto finito de índices C. donde cada conjunto X_i es no vacío, cerrado y convexo. Además, asumimos que la función u es separable con respecto a la partición de $\mathbf{R}^n$ definida en (1.9), esto es, u es de la forma

$$u(x) = \sum_{i \in C} u_i(x_i),$$

donde $u_i : \mathbf{R}^{n_i} \mapsto \mathbf{R} \cup \{+\infty\}$ es una función convexa, propia y lsc para cada $i \in C$. El mapeo F en general es no separable con respecto a la partición dada de $\mathbf{R}^n$; de lo contrario, el problema GVIP(F, u, X) podría ser descompuesto en un número de problemas independientes de la forma GVIP(F_i, u_i, X_i). (Por esto podemos argumentar que el problema dado generaliza GVIP(F, u, X))

Ejemplo 1.3 (Problema del Equilibrio de Nash). Sea $X = \prod_{i=1}^{N} X_i$ el producto de conjuntos individuales de estrategia convexos, cerrados y no vacíos $X_i \subseteq \mathbf{R}^{n_i}$, $\sum_{i=1}^{N} n_i = n$. Definimos una función de penalidad convexa y de clase C^1 $f_i : X \mapsto \mathbf{R}$ para cada jugador definida en el espacio unido de estrategias. Además, definimos la función $x \mapsto u(x) = \sum_{i=1}^{N} u_i(x_i)$ separable, convexa, propia y lsc en X. Un equilibrio de Nash del juego no-cooperativo de N-personas con estos datos es descrito por el punto $x^* \in X$ con el que, para cada $i \in \{1, \ldots, N\}$, satisface

$$f_i(x_j^*, x_i^*) + u_i(x_i^*) = \underset{x_i \in X_i}{\text{minimizar}} \{f_i(x_j^*, x_i) + u_i(x_i)\}, \quad i \neq j \tag{1.10}$$

lo que quiere decir que las estrategias de los jugadores son óptimas respecto a sus funciones individuales de penalidad, basadas en las estrategias de los otros jugadores. Las condiciones de óptimo para (1.10) definen una instancia de GVIP(F, u, X) de la forma descrita en el Ejemplo 1.5, y en este caso tenemos $x = \{x_1, \ldots, x_N\}$, $X = \prod_{i=1}^{N} X_i$ y $F = (\nabla_{x_1} f_1, \ldots, \nabla_{x_N} f_N)$.

<u>*Observación*</u> 1.4 (No existe única representación para GVIP(F, u, X)). GVIP(F, u, X) no está definida únicamente en términos de la tri-upla $[F, \partial u, N_X]$. Por ejemplo, el conjunto X puede ser representado añadiendo la función indicadora δ_X de X a u. Esta función de penalidad infinita es propia, convexa y lsc, así como lo es $u + \delta_X$, se cumple que $\partial\delta_X = N_X$, bajo la condición de regularidad 1.1, $\partial[u + \delta_X](x) = \partial u(x) + N_X(x)$. Así entonces, cualquiera que sea la restricción convexa, puede ser reemplazada o en la descripción de X o como una penalidad infinita agregada a la descripción de u y considerando que se cumple la condición 1.1, entonces no habría ninguna pérdida de generalidad en expresar GVIP(F, u, X) como la *Ecuación Generalizada*

$$0^n \in F(x^*) + \partial u(x^*) \tag{1.11}$$

al cual vamos a denotar como GE(F, u), que es como si hiciéramos $u = u + \delta_X$.

Por otra parte, la descomposición de F y ∂u tampoco es única, por ejm., podríamos agregar y sustraer el gradiente ∇h de una función convexa arbitraria h a F y a ∂u, lo cual dejaría GVIP(F, u, X) inalterado.

Entonces, debido a la descomposición no única de N_X y ∂u; y de, F y ∂u, hay una gran libertad de elección en representar una instancia de GVIP(F, u, X) en términos de estos mapeos, lo cual es de gran utilidad a la hora de usar algoritmos para resolver estos tipos de problemas, ya que podemos cambiar la representación que GVIP(F, u, X), de modo que concuerde con los requerimientos de aproximación por parte de los mapeos y convergencia a la solución que nos brinda el propio algoritmo. ■

Si $u = 0$ en GVIP(F, u, X) (notar que la condición de regularidad se ve satisfecha trivialmente), lo cual es equivalente a asumir que en GE(F, u), $u = \delta_X$ para algún conjunto convexo, cerrado y no vacío $X \subseteq \mathbb{R}^n$, entonces obtenemos el *Problema de Desigualdad Variacional* (VIP(F, X)) de encontrar $x^* \in \mathbb{R}^n$ tal que

$$0^n \in F(x^*) + N_X(x^*), \tag{1.12}$$

o como es más conocido (usando (1.1)), hallar $x^* \in X$ tal que

$$F(x^*)^T(x - x^*) \geq 0, \qquad \forall x \in X \tag{1.13}$$

<u>*Observación*</u> 1.5 (Caracterización de Proyección). Tomemos un $\gamma > 0$. Podemos entonces escribir (1.13) de la siguiente manera

$$[(1/\gamma)x^* + F(x^*) - (1/\gamma)x^*]^T(x - x^*) \geq 0, \qquad \forall x \in X$$

$$\Rightarrow$$

$$[(x^* - \gamma F(x^*)) - x^*]^T(x - x^*) \leq 0, \qquad \forall x \in X;$$

esta desigualdad muestra que $x^* - \gamma F(x^*) - x^*$ pertenece al cono normal de X en x^*, es decir, x^* es la Proyección Euclidiana del vector $x^* - \gamma F(x^*)$ sobre X. Entonces $x^* = P_X[x^* - \gamma F(x^*)]$.

De manera más general, sea $Q \in \mathbf{R}^{n\times n}$ una matriz definida positiva y simétrica. Trabajando análogamente obtenemos que una solución x^* para VIP(F, X) es caracterizada por la desigualdad

$$[(x^* - \gamma Q^{-1}F(x^*)) - x^*]^T Q(x - x^*) \leq 0, \qquad \forall x \in X,$$

que es equivalente a decir que x^* es la proyección del vector $x^* - \gamma Q^{-1}F(x^*)$ sobre X usando la norma $\|z\|_Q = \sqrt{z^T Q z}$ definida por Q, esto es, $x^* = P_X^Q[x^* - \gamma Q^{-1}F(x^*)]$. ■

Ejemplo 1.6 (Equilibrio de Transporte). Sea G = (N,A) que denota una red de transporte urbano con nodos (intersecciones y centroides) y de conexiones directas (caminos). Un subconjunto C de $N \times N$ define un conjunto de bienes en el que cada elemento representa un camino origen-llegada. Se asume que la demanda de transporte d_k en cualquiera de los caminos k de C es conocida. Haciendo x_{k_r}, $r \in R_k$, el flujo en la ruta r perteneciente al conjunto de rutas R_k para el bien k, el conjunto de las rutas factibles es descrito por las restricciones

$$\text{(Satisfaccion de la demanda)} \sum_{r\in R_k} x_{k_r} = d_k, \qquad \forall k \in C, \tag{1.14a}$$

$$\text{(No negatividad)} \qquad x_{k_r} \geq 0, \qquad \forall k \in C, \tag{1.14b}$$

o de forma compacta

$$x = \{x \in \mathbf{R}^{|R|} \mid \Gamma^T x = d; \quad x \geq 0^{|R|}\}$$

donde $R = \cup_k R_k$ y $\Gamma = (\gamma_{kr})$ es la matriz de incidencia de pares de ruta-OD tales que

$$\gamma_{kr} = \begin{cases} 1, & \text{si la ruta } r_k \text{ une el par OD } k, \\ 0, & \text{en otro caso,} \end{cases} \qquad r \in R_k, \quad k \in C$$

Vamos a asumir además que cada ruta $r \in R_k$, $k \in C$, está asociada con una función costo de ruta F_{kr}: $\mathbf{R}_+^{|R|} \mapsto \mathbf{R}_{++}$, que mide la desutilidad de viajar por dicha ruta en función del volumen de trafico en la red. Bajo la condición de que el viajero escoge la ruta para su destino de menor costo dadas las condiciones actuales de la red, el equilibrio se obtiene cuando ningún viajero puede reducir su costo de viaje cambiando de ruta. Por lo que, todas las rutas usadas entre un par OD dado tienen el mismo costo mínimo. Este equilibrio se describe como

$$x_{kr}[F_{kr}(x) - \pi_k] = 0, \qquad r \in R_k, \quad k \in C \tag{1.15a}$$

$$F_{kr}(x) - \pi_k \geq 0, \qquad r \in R_k, \quad k \in C \tag{1.15b}$$

(donde π_k es el costo mínimo de viaje en el par k-ésimo OD).
(1.15a) Dice que si el costo de viajar por la ruta r en el par k-ésimo OD es mayor al costo mínimo, entonces el flujo correspondiente x_{kr} es cero, esto es, dicha ruta no debería llevar flujo, a veces se dice que por una ruta cara no hay intensión de viajar; y si x_{kr} es positivo, es porque el costo es barato, esto es, igual al costo mínimo.
Estas condiciones para el flujo de ruta factible son equivalentes al VIP(F, X), con $F = (F_{kr})_{r \in R_k, k \in C}$, que resultar ser una instancia para el problema descrito en el ejemplo (1.5). ∎

Ejemplo 1.7 (Problema de punto silla). Sean $V \subseteq \mathbb{R}^n$ y $W \subseteq \mathbb{R}^m$ conjuntos convexos cerrados, y $Z : V \times W \mapsto \mathbb{R}$ una función continua en $V \times W$. El problema de punto silla asociado con Z es hallar un $(v^*, w^*) \in V \times W$ tal que

$$Z(v^*, w) \leq Z(v^*, w^*) \leq Z(v, w^*), \qquad \forall (v, w) \in V \times W \tag{1.16}$$

Entonces, condiciones necesarias para el problema de punto silla en (v^*, w^*) son

$$0^m \in -\nabla_w Z(v^*, w^*) + N_W(w^*); \qquad 0^n \in \nabla_v Z(v^*, w^*) + N_V(v^*) \tag{1.17}$$

Haciendo $x = (v, w)$, $X = V \times W$, y $F = (\nabla_v Z, -\nabla_w Z)$, otra vez estamos en un caso especial del problema del ejemplo (1.5). ∎

Asumamos que F es el gradiente de una función continuamente diferenciable $f : \operatorname{dom} u \cap X \mapsto \mathbb{R}$ (f es de clase C^1 en $\operatorname{dom} u \cap X$)Entonces, (1.4) define las condiciones necesarias de óptimo para el *Programa No Diferenciable Restringido* [CNDP(f, u, X)]

$$\underset{x \in X}{\text{minimizar}}\ T(x) = f(x) + u(x) \tag{1.18}$$

y otra vez, sin pérdida de generalidad, como no hemos asumido que el dom $u = \mathbb{R}^n$, incluyendo δ_X en u, podemos escribir esto como el siguiente *Programa No Diferenciable no Restringido*

[NDP(f, u)]

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}}\ T(x) = f(x) + u(x) \tag{1.19}$$

<u>*Observación*</u> 1.8 (Propiedades de Diferenciabilidad de T). Como T es la suma de una función diferenciable y una convexa, entonces resulta ser localmente Lipschitz continua, lo cual es suficiente para poder decir que es ***semisuave*** y ***quasi-diferenciable***. ■

Ahora, si asumimos que $u = 0$, VIP(F, X) define las condiciones de primer orden necesarias de óptimo para el ***Programa Diferenciable Restringido*** [CDP(f, X)]

$$\underset{x \in X}{\text{minimizar}}\ f(x) \tag{1.20}$$

Vamos a denotar al conjunto de soluciones globales de CDP(f, X) como SOL(f, X), viendo que SOL(f, X) $\subseteq$ SOL($\nabla f, X$).

Ejemplo 1.9 (Reformulación de penalidad de CDP(f, X)).
Consideremos el problema CDP($f, X \cap G$), con las restricciones implícitas

$$G = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, \quad i = 1, \ldots, m\},$$

donde $g_i : X \mapsto \mathbb{R}^n$, $i = 1, \ldots, m$, son convexas en X. Vamos a tratar de reemplazar estas restricciones por una función de penalidad. Sea $p : \mathbb{R} \mapsto \mathbb{R}$ una función creciente, no-negativa y convexa tal que $p(z) = 0$ si y solo si, $z \leq 0$ (el ejemplo más conocido de este tipo de funciones es $p(z) = \text{máx}\{0, z\}$). Asi, sea $r > 0$, entonces una reformulación de penalidad de CDP($f, X \cap G$) es el problema de optimización

$$\underset{x \in X}{\text{minimizar}}\ f(x) + r \sum_{i=1}^{m} p(g_i(x)).$$

Notar que esta reformulación es una instancia de CNDP(f, u, X). ■

Si se toma $u = 0$, GE(F, u) se convierte en el siguiente sistema de ecuaciones (posiblemente no lineales) [NLE(F)]

$$F(x^*) = 0^n; \tag{1.21}$$

siempre que F sea la gradiente de una función f, este sistema define las condiciones de óptimo de primero orden para el ***Programa No Lineal*** [NLP(f)]

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}}\ f(x) \tag{1.22}$$

Capítulo 2

Formulaciones Especiales

En esta parte desarrollaremos algunos aspectos que ilustran que los GVIP(F, u, X) y sus instancias incluyen formulaciones que involucran tanto las cantidades primales como las duales.

Formulación 2.1 (Desigualdades Variacionales Primal-Dual). consideremos el siguiente caso especial de GVIP(F, u, X)

$$0^n \in H(y^*) + N_{Y\cap G}(y^*) \tag{2.1}$$

donde $Y \subseteq \mathbb{R}^n$ es un conjunto no vacío, cerrado y convexo; $H : \mathbb{R}^n \mapsto \mathbb{R}$ es un función continua y $G = \{y \in \mathbb{R}^n \,|\, g_i(y) \leq 0, \quad i = 1, \ldots, m\}$, donde cada función $g_i : \mathbb{R}^n \mapsto \mathbb{R}$ es convexa y de clase C^1. Introduciendo multiplicadores $\lambda_i \in \mathbb{R}_+$, $i = 1, \ldots, m$ para las restricciones que definen G, resulta la siguiente formulación primal-dual:

$$0^{n+m} \in \begin{pmatrix} H(y^*) + \nabla g(y^*)^T \lambda^* \\ -g(y^*) \end{pmatrix} + N_{Y\times\mathbb{R}^m_+}(y^*, \lambda^*) \tag{2.2}$$

Esta desigualdad variacional en las variables (y, λ) describe las condiciones de Karush-Kuhn-Tucker (KKT) para (2.1); y bajo una apropiada restricción de calificación (CQ), las soluciones en y coinciden.

Identificamos a (2.2) como un caso especial del problema del ejemplo 1.5, si hacemos $x = (y, \lambda)$, $X = Y \times \mathbb{R}^m_+$, y $F(x) = [H(y) + \nabla g(y)^T \lambda, -g(y)]$ ∎

Formulación 2.2 (Condiciones de Karush-Kuhn-Tucker). Consideremos el siguiente problema de optimización (posiblemente no-convexo)

$$\text{minimizar} \quad f(y) \tag{2.3a}$$
$$\text{sujeto a} \quad g(y) \in K^o, \tag{2.3b}$$
$$y \in Y, \tag{2.3c}$$

donde f y g son funciones de clase C^1 que van de algún conjunto abierto de $\mathbb{R}^p$ a $\mathbb{R}$ y a $\mathbb{R}^m$ respectivamente, Y es un conjunto convexo y cerrado en $\mathbb{R}^p$, K es un cono convexo cerrado en $\mathbb{R}^m$, y $K^o = \{z \in \mathbb{R}^m \mid z^T s \leq 0, \quad \forall s \in K\}$ es el *Cono Polar* de K. Por ejemplo, si las restricciones son de la forma

$$g_i(y) \leq 0, \quad i = 1, \ldots, k; \qquad g_i(y) = 0, \quad i = k+1, \ldots, k+l, \tag{2.4}$$

entonces $K = \mathbb{R}^k_+ \times \mathbb{R}^l$ y $K^o = \mathbb{R}^k_- \times \{0\}^l$. La función Lagrangeana asociada con (2.3b) es $L(y, \lambda) = f(y) + \lambda^T g(y)$, $y \in Y$, $\lambda \in \mathbb{R}^m$; bajo apropiadas CQ, cualquier solución óptima local y para (2.3) satisfará las condiciones KKT necesarias, que pueden ser vistas como una instancia del VIP(F, X), con $x = (y, \lambda)$, $X = Y \times K$, y $F(x) = [\nabla_y L(y, \lambda), -\nabla_\lambda L(y, \lambda)]$ ∎

Formulación 2.3 (Problemas de Complementariedad). Dados una función $F : \mathbb{R}^n \mapsto \mathbb{R}^n$ y un cono convexo cerrado $K \subset \mathbb{R}^n$, el *Problema Generalizado de Complementariedad* [GCP(F, K)] es encontrar un $x \in \mathbb{R}^n$ tal que

$$x \in K; \qquad F(x) \in K^*; \qquad x^T F(x) = 0, \tag{2.5}$$

donde $K^* = -K^o$ es el cono dual de K.

Notar que (2.5) se puede ver como un caso muy especial de un VIP(F, X) para $X = K$. Por otro lado, consideremos el problema de desigualdad variacional VIP(H, G), con un mapeo continuo $H : \mathbb{R}^n \mapsto R^n$ y con restricciones explícitas donde G es dado por (2.4), y cada $g_i : \mathbb{R}^n \mapsto \mathbb{R}$ es convexa y de clase C^1. Introduciendo vectores multiplicadores $\lambda \in \mathbb{R}^k_+ \times \mathbb{R}^l$ para las restricciones, bajo un CQ, VIP(H, G) es equivalente a un GCP(F, K) donde

$$x = (y, \lambda); \qquad K = \mathbb{R}^n \times \mathbb{R}^k_+ \times \mathbb{R}^l; \qquad F(x) = \begin{pmatrix} H(y) + \nabla g(y)^T \lambda \\ -g(y) \end{pmatrix}$$

Este GCP extiende las condiciones KKT para el problema de optimización no lineal del ejemplo 2.3 en el caso cuando $Y = \mathbb{R}^n$ y $K = \mathbb{R}^k_+ \times \mathbb{R}^l$, y además, se reduce a una instancia de NLE(F) cuando k = 0. Mas aún, es un caso muy especial del problema (2.2) para el caso cuando $Y = \mathbb{R}^n$. El *Problema de Complementariedad No Lineal* [NCP(F)]

$$x \geq 0^n; \qquad F(x) \geq 0^n; \qquad x^T F(x) = 0. \tag{2.6}$$

se obtiene si $K = \mathbb{R}^n_+$; y si además F es afín, es decir, es de la forma $F : x \mapsto F(x) = Ax - b$ para alguna matriz $A \in \mathbb{R}^{n \times n}$ y un vector $b \in \mathbb{R}^n$, entonces el problema (2.6) es un *Problema de Complementariedad Lineal* (LCP). ∎

Formulación 2.4 (Proyección / Particionamiento). Considere el problema

$$\begin{aligned} \text{minimizar} \quad & f(x) + h(y) \\ \text{sujeto a} \quad & c_i(x) + d_i(y) \le b_i, \qquad i = 1, \dots, p, \\ & x \in X, y \in Y, \end{aligned} \tag{2.7}$$

donde $X \subseteq \mathbf{R}^n$ y $Y \subseteq \mathbf{R}^m$ son conjuntos no vacíos, cerrados y convexos, $f : \mathbf{R}^n \mapsto \mathbf{R}$ es de clase C^1, $h : \mathbf{R}^m \mapsto \mathbf{R}$, $c_i : \mathbf{R}^n \mapsto \mathbf{R}$, y $d_i : \mathbf{R}^m \mapsto \mathbf{R}$, $i = 1, \dots, p$, son funciones convexas, y $b \in \mathbf{R}^p$. Vamos a resolver este problema en términos de x eliminando las variables y, esto es, expresando la solución óptima en y para cada valor fijo x. Reescribiendo (2.7) como corresponde, entonces, obtenemos una instancia de CNDP(f, u, X), donde

$$\begin{aligned} u(x) = \text{infimo} \quad & h(y) \\ \text{sujeto a} \quad & d_i(y) \le b_i - c_i(x), \qquad i = 1, \dots, p, \\ & y \in Y, \end{aligned} \tag{2.8}$$

Notemos que u es convexo, y que su dominio es el subconjunto de X para el cual el ínfimo en (2.8) es finitamente obtenido. ∎

Formulación 2.5 (Descomposición Dual). Un caso especial de la formulación proyección/particionamiento (2.8) surge cuando todas las funciones involucradas en (2.7) son lineales, y las restricciones poseen la estructura dual de bloque como se ve abajo. Consideremos el programa lineal

$$\begin{aligned} \text{minimizar} \quad & v^T x + w^T y, \\ \text{sujeto a} \quad & Ax + By = b, \\ & Cx \qquad = c, \\ & \qquad Dy = d, \\ & x, \quad y \le 0^{n+m}, \end{aligned}$$

donde $A \in \mathbf{R}^{p \times n}$, $B \in \mathbf{R}^{p \times m}$, $C \in \mathbf{R}^{q \times n}$, $D \in \mathbf{R}^{r \times m}$, y los vectores v, w, b, c y d tienen las dimensiones apropiadas. Si asumimos que el conjunto $V = \{v \in \mathbf{R}^{p+r} \mid B^T v_1 + D^T v_2 \le w\}$ es no vacío, y denotamos como v^j, $j \in \mathcal{P}$ (v^i, $i \in \mathcal{D}$, respectivamente) el conjunto de sus puntos extremos (direcciones), entonces el problema anterior es equivalente a

$$\begin{aligned} \text{minimizar} \quad & v^T x + u(x), \\ \text{sujeto a} \quad & Cx = c, \\ & x \ge 0^n, \\ & (b - Ax)^T v_1^i + w^T v_2^i \le 0, \qquad i \in \mathcal{D}, \end{aligned}$$

donde

$$\begin{aligned} u(x) = \text{minimizar} \quad & w^T y, \\ \text{sujeto a} \quad & By = b - Ax, \\ & Dy = d, \\ & y \geq 0^m, \end{aligned}$$

■

Formulación 2.6 (Particionamiento como representación de variables no básicas). Consideremos el problema

$$\begin{aligned} \text{minimizar} \quad & f(x), && (2.9a) \\ \text{sujeto a} \quad & Ax = b, && (2.9b) \\ & x \geq 0^n, && (2.9c) \end{aligned}$$

donde $A \in \mathbb{R}^{m \times n}$ tiene rango de fila completo y $n > m$. Asumamos que $x = (x_B^T, x_N^T)^T$ y $A = (B, N)$ define una partición de $\mathbb{R}^n$ tal que $B \in \mathbb{R}^{m \times m}$ es invertible y $B^{-1}b \geq 0^m$. Como x_B puede ser representado, únicamente, de la forma $x_B = B^{-1}[b - Nx_N]$, podemos escribir (2.9) en términos de las variables no básicas x_N como

$$\begin{aligned} \text{minimizar} \quad & f(B^{-1}[b - Nx_N], x_N), \\ \text{sujeto a} \quad & x_N \geq 0^{n-m}, \\ & B^{-1}[b - Nx_N] \geq 0^m, \end{aligned}$$

■

Capítulo 3

Teoría de Optimización No Diferenciable

En este capítulo vamos a presentar algunos resultados que conectan las teorías de análisis no suaves y optimización. Primero generalizaremos las condiciones clásicas de óptimo tanto para el caso restringido como para el no restringido. Daremos las condiciones necesarias para que una función de Lipschitz alcance su mínimo local. Para funciones convexas, estas condiciones son también suficientes y el mínimo es global.

3.1. Definiciones y Teoremas

Vamos a dar algunas definiciones y notaciones que usaremos para nuestros resultados:

Definición 3.1 (Derivada Direccional). La derivada direccional de la función f en la dirección $v \in \mathbf{R}^n$ se define como

$$f'(x; v) = \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t} \tag{3.1}$$

Definición 3.2 (Subdiferencial). El *Subdiferencial* de una función convexa $f : \mathbf{R}^n \to \mathbf{R}$ en $x \in \mathbf{R}^n$ es el conjunto

$$\partial_c f(x) = \{\xi \in \mathbf{R}^n \mid f(y) \geq f(x) + \langle \xi, y - x \rangle, \quad \forall y \in \mathbb{R}^n\} \tag{3.2}$$

Cada elemento $\xi \in \partial_c f(x)$ es llamado *subgradiente* de f en x.

Teorema 3.3. Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Entonces para cada x tenemos

(i) $f'(x; v) = \max\{\langle \xi, v \rangle \mid \xi \in \partial_c f(x)\}$, para todo $v \in \mathbf{R}^n$

(ii) $\partial_c f(x) = \{\xi \in \mathbf{R}^n \mid f'(x;v) \geq \langle \xi, v\rangle$, para todo $v \in \mathbf{R}^n\}$

(iii) $\partial_c f(x)$ es un conjunto no vacío, convexo y compacto tal que $\|\xi\| \leq K$ para todo $\xi \in \partial_c f(x)$, donde K es la constante local de Lipschitz de f en x,

(iv) el mapeo punto-conjunto $\partial_c f(\cdot) : \mathbf{R}^n \to \mathcal{P}(\mathbf{R}^n)$ es semicontinua superiormente, i.e., si $y_i \to x$ y $\xi_i \in \partial_c f(y_i)$ para cada i, entonces cada punto de acumulación ξ de (ξ_i) está en $\partial_c f(x)$.

Teorema 3.4. Si $f : \mathbf{R}^n \to \mathbf{R}$ es convexa y diferenciable en $x \in \mathbf{R}^n$, entonces

$$\partial_c f(x) = \{\nabla f(x)\} \tag{3.3}$$

Teorema 3.5. Si $f : \mathbf{R}^n \to \mathbf{R}$ es convexa, entonces para todo $y \in \mathbf{R}^n$

$$f(y) = \max\{f(x) + \langle \xi, y - x\rangle \mid x \in \mathbf{R}^n,\ \xi \in \partial_c f(x)\} \tag{3.4}$$

Definición 3.6 (Cono Tangente). Sea C un conjunto no vacío convexo de $\mathbf{R}^n$. El *Cono Tangente* de C en el punto $x \in C$ es el conjunto

$$T_C(x) = \{y \in \mathbf{R}^n \mid \text{existe } t_i \downarrow 0 \quad \text{y} \quad y_i \to y \quad \text{con} \quad x + t_i y_i \in C\} \tag{3.5}$$

Definición 3.7 (Cono Normal). Sea C un conjunto no vacío convexo de $\mathbf{R}^n$. El *Cono Normal* de C en el punto $x \in C$ es el conjunto

$$N_C(x) = T_C(x)^\circ = \{z \in \mathbf{R}^n \mid \langle y, z\rangle \leq 0 \quad \forall y \in T_C(x)\} \tag{3.6}$$

Definición 3.8. Sea $G \subset \mathbf{R}^n$ un conjunto no vacío. La *función distancia* $d_G : \mathbf{R}^n \to \mathbf{R}$ para el conjunto G es definida por

$$d_G(x) = \inf\{\|x - c\|, \quad c \in G\} \tag{3.7}$$

Definición 3.9 (Epígrafo). El *epígrafo* de una función $f : \mathbf{R}^n \to \mathbf{R}$ es el siguiente subconjunto de $\mathbf{R}^n \times \mathbf{R}$

$$\operatorname{epi} f = \{(x, r) \in \mathbf{R}^n \times \mathbf{R} \mid f(x) \leq r\} \tag{3.8}$$

Definición 3.10 (Derivada Direccional Generalizada). Sea $f : \mathbf{R}^n \mapsto \mathbf{R}$ localmente Lipschitz en un punto $x \in \mathbf{R}^n$. La *Derivada Direccional Generalizada* de f en x en la dirección $v \in \mathbf{R}^n$ está definida como

$$f^\circ(x;v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t} \tag{3.9}$$

Definición 3.11. Sea $f : \mathbf{R}^n \to \mathbf{R}$ localmente Lipschitz en $x \in \mathbf{R}^n$. Entonces el *Subdiferencial* de f en x es el conjunto

$$\partial f(x) = \{\xi \in \mathbf{R}^n \mid f^{\circ}(x; v) \geq \langle \xi, v \rangle \quad \text{para todo} \quad v \in \mathbf{R}^n\} \tag{3.10}$$

Cada elemento $\xi \in \partial f(x)$ es llamado ***subgrdiente*** de f en x.

Definición 3.12 (Función Regular). La función $f : \mathbf{R}^n \to \mathbf{R}$ se dice que es *Regular* en $x \in \mathbf{R}^n$ si para todo $v \in \mathbf{R}^n$ la derivada direccional $f'(x; v)$ existe y

$$f'(x; v) = f^{\circ}(x; v) \tag{3.11}$$

Teorema 3.13 (Regularidad). Sea f Lipschitz en x. Entonces f es regular en x si:

i.- f es continuamente diferenciable en x

ii.- f es convexa

iii.- $f = \sum_{i=1}^{m} \lambda_i f_i$, donde $\lambda_i > 0$ y f_i es regular para cada $i = 1, \ldots, m$

Dadas estas definiciones, vamos a nombrar algunos teoremas que nos serán útiles en nuestro análisis:

3.2. Teoremas Básicos

Teorema 3.14 (Hahn-Banach). Sea X un espacio vectorial lineal real y $Y \subset X$ un subespacio lineal. Si $p : X \to \mathbf{R}$ es una funcional positivamente homogénea, subaditiva y $f : Y \to \mathbf{R}$ una funcional lineal tal que $f(y) \leq p(y)$ para todo $y \in Y$, entonces existe una funcional lineal $F : X \to \mathbf{R}$ tal que

$$F(y) = f(y) \quad \text{para todo} \quad y \in Y \quad \text{y} \quad F(x) \leq p(x) \quad \text{para todo} \quad x \in X \tag{3.12}$$

Prueba.- [Rudin-1973, pág. 56] ∎

Teorema 3.15. La cápsula convexa de un conjunto compacto es compacta

Prueba.- [Rudin-1973] ∎

Teorema 3.16. Sean C y D conjuntos convexos y compactos en $\mathbf{R}^n$. Entonces $C \subset D$ sii

$$\max_{c \in C} \langle c, x \rangle \leq \max_{d \in D} \langle d, x \rangle \quad \text{para todo} \quad x \in \mathbf{R}^n \tag{3.13}$$

Prueba.- [Rockafellar-1970, pág. 113] ∎

Teorema 3.17. Sea U un conjunto abierto en $\mathbb{R}^n$ y $f : U \to \mathbb{R}$. Entonces f es diferenciable en casi todos lados en U si

$$\limsup_{t \downarrow 0} \left| \frac{f(x+tv) - f(x)}{t} \right| < \infty \quad \text{para todo} \quad x, v \in \mathbb{R}^n \tag{3.14}$$

<u>*Prueba*</u> .- [Cohn-1980] ∎

Teorema 3.18 (Bolzano-Weierstrass). Si un conjunto $C \in \mathbb{R}^n$ contiene infinitos puntos, entonces existe al menos un punto en $\mathbb{R}^n$ que es punto de acumulación de C.

<u>*Prueba*</u> .- [EuJa-2000] ∎

Teorema 3.19 (Fubini). Si f es una función medible sobre el producto cartesiano $X \times Y$ de dos espacios vectoriales reales y si $\int\int |f| dudv < \infty$ o $\int\int |f| dvdu < \infty$, entonces f es integrable en $X \times Y$ y

$$\int\int f dudv = \int\int f dvdu \tag{3.15}$$

<u>*Prueba*</u> .- [Cohn-1980] ∎

3.3. Análisis Subdiferencial

Teorema 3.20 (Valor Medio). Sean $x, y \in \mathbb{R}^n$ y sea la función f de Lipschitz en el conjunto abierto $U \subset \mathbb{R}^n$ tal que el segmento de linea $[x, y] \subset U$. Entonces existe un punto $u \in [x, y]$ tal que

$$f(y) - f(x) \in \langle \partial f(u), y - x \rangle \tag{3.16}$$

Para esta prueba vamos a necesitar del siguiente Lema:

Lema 3.21. Sea la función f de Lipschitz, entonces la función $g : [0,1] \to \mathbb{R}$ definida por $g(t) = f(x + t(y-x))$ es de Lipschitz en $\langle 0, 1 \rangle$ y

$$\partial g(t) \subset \langle \partial f(x + t(y-x)), y - x \rangle \tag{3.17}$$

<u>*Prueba*</u> .- Denotemos a $x + t(y-x)$ por x_t. Entonces

$$\begin{aligned} |g(t) - g(t')| &= |f(x_t) - f(x_{t'})| \leq K \|x_t - x_{t'}\| \\ &= K \|(t-t')(y-x)\| = K\|y-x\| |t-t'| \\ &= \tilde{K} |t - t'| \quad \text{para todo} \quad t, t' \in \langle 0, 1 \rangle \end{aligned}$$

donde $\tilde{K} = K\|y - x\|$

Ahora, como las funciones f y g son Lipschitz, los conjuntos $\partial g(t)$ y $\langle \partial f(x_t), y-x\rangle$ son compactos y convexos. Como estos conjuntos pertenecen a $\mathbb{R}$, deben ser intervalos cerrados en $\mathbb{R}$, entonces para obtener la demostración del Lema es suficiente con demostrar que para $v = \pm 1$, tenemos

$$\text{máx}\{\partial g(t)v\} \leq \text{máx}\{\langle \partial f(x_t), y - x\rangle v\}$$

Sabemos que $\text{máx}\{\partial g(t)v\} = g^{\circ}(t; v)$ (tambien se cumple para la derivada direccional generalizada, ver teorema 3.3) por lo que

$$\begin{aligned}
\text{máx}\{\partial g(t)v\} &= \limsup_{\substack{s\to t\\ \lambda\downarrow 0}} \frac{g(s+\lambda v) - g(s)}{\lambda} \\
&= \limsup_{\substack{s\to t\\ \lambda\downarrow 0}} \frac{f(x + (s+\lambda v)(y-x)) - f(x + s(y-x))}{\lambda} \\
&\leq \limsup_{\substack{y'\to x_t\\ \lambda\downarrow 0}} \frac{f(y' + \lambda v(y-x)) - f(y')}{\lambda} \\
&= f^{\circ}(x_t; v(y-x))
\end{aligned}$$

Y usando el criterio anterior,

$$f^{\circ}(x_t; v(y-x)) = \text{máx}\{\langle \partial f(x_t), v(y-x)\rangle\}$$

por lo que

$$\text{máx}\{\partial g(t)v\} \leq \text{máx}\{\langle \partial f(x_t), y - x\rangle v\}$$

■

Demostremos ahora el teorema del valor medio:

Prueba.- **(Valor Medio)**

Definamos la función $\Theta : [0,1] \to \mathbb{R}$ tal que $\Theta(t) = f(x_t) + t(f(x) - f(y))$, donde $x_t = x + t(y-x)$. Entonces, al ser Θ una función continua y

$$\Theta(0) = f(x_0) = f(x)$$

$$\Theta(1) = f(x_1) + f(x) - f(y) = f(x)$$

Entonces existe un $t_0 \in \langle 0, 1\rangle$ tal que $\Theta(t)$ alcanza un valor extremo en t_0, por lo que $0 \in \partial\Theta(t_0)$ y así

$$\partial\Theta(t) = \partial[f(x_t) + t(f(x) - f(y))] \subset \partial f(x_t) + [f(x) - f(y)]\partial(t)$$

y además, por el lema 3.21 tenemos

$$0 \in \langle \partial f(x_t), y - x\rangle + [f(x) - f(y)]\partial(t)$$

y como $\partial(t) = 1$ se sigue que

$$f(y) - f(x) \in \langle \partial f(u), y - x\rangle$$

donde $u = x_{t_0} \in \langle x, y\rangle$, lo que demuestra el teorema. ■

Teorema 3.22 (Regla de la Cadena). Sean $h : \mathbb{R}^n \to \mathbb{R}^m$ y $g : \mathbb{R}^m \to \mathbb{R}$ funciones tales que cada función componente $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ es localmente Lipschitz en $x \in \mathbb{R}^n$ y g es localmente Lipschitz en $h(x) \in \mathbb{R}^m$. Entonces la función composición $f = g \circ h$, $f : \mathbb{R}^n \to \mathbb{R}$ es localmente Lipschitz en x y

$$\partial f(x) \subset \text{conv}\left\{\sum_{i=1}^{m} \alpha_i \xi_i \mid \xi_i \in \partial h_i(x) \quad \text{y} \quad \alpha \in \partial g(h(x))\right\} \tag{3.18}$$

Donde la igualdad se da bajo cualquiera de las siguientes hipótesis:

(i) La función g es regular en $h(x)$, cada h_i es regular en x y $\alpha_i \geq 0$ para todo $i = 1, \ldots, m$. Entonces f también es regular en x.

(ii) La función g es regular en $h(x)$ y h_i es continuamente diferenciable en x para todo $i = 1, \ldots, m$

(iii) Si $m = 1$ y g es continuamente diferenciable en $h(x)$.

Prueba.- Dada la definición de f, se deduce que es localmente Lipschitz en x. Definamos

$$S = \{\sum_{i=1}^{m} \alpha_i \xi_i \mid \xi_i \in \partial h_i(x) \quad \text{y} \quad \alpha \in \partial g(h(x))\}$$

Como $\partial h_i(x)$ y $\partial g(h(x))$ son conjuntos compactos, entonces el conjunto S también es compacto, y por el teorema 3.15 su capsula convexa es un conjunto convexo compacto. Entonces por el teorema 3.16 y de la definición de $\partial f(x)$ es suficiente con mostrar que

$$f^\circ(x; v) \leq \max_{\eta \in \text{conv } S} \langle \eta, v\rangle \quad \text{para todo} \quad v \in \mathbb{R}^n \tag{3.19}$$

Para ver esto, supongamos que $\eta \in \text{conv } S$. Entonces $\eta = \sum_{j=1}^{k} \lambda^j s^j$ con $s^j \in S$ y $\sum_{j=1}^{k} \lambda^j = 1$ con $\lambda_j \geq 0$. Entonces para todo $v \in \mathbb{R}^n$ tenemos

$$\langle \eta, v\rangle = \sum_{j=1}^{k} \lambda^j \langle s^j, v\rangle \leq \sum_{j=1}^{k} \lambda^j \max_{s \in S}\langle s, v\rangle = \max_{s \in S}\langle s, v\rangle$$

entonces

$$\max_{\eta \in \operatorname{conv} S} \langle \eta, v\rangle = \max_{s \in S} \langle s, v\rangle \quad \text{para todo} \quad v \in \mathbb{R}^n$$

Definamos ahora

$$q_\varepsilon(v) = \max \Big\{ \sum_{i=1}^{m} \alpha_i \langle \xi_i, v\rangle \mid \xi_i \in \partial h_i(x_i), \alpha \in \partial g(u), x_i \in B(x,\varepsilon), u \in B(h(x),\varepsilon) \Big\}$$

Entonces tenemos

$$\begin{aligned} q_0(v) &= \max \Big\{ \sum_{i=1}^{m} \alpha_i \langle \xi_i, v\rangle \mid \xi_i \in \partial h_i(x), \alpha \in \partial g(h(x)) \Big\} \\ &= \max_{s \in S} \langle s, v\rangle \end{aligned}$$

Esto implicaría (3.19) si mostramos que para $\varepsilon > 0$

$$f^{\circ}(x; v) - \varepsilon \leq q_\varepsilon(v) \tag{3.20}$$

y que $q_\varepsilon(v) \to q_0(v)$ mientras $\varepsilon \downarrow 0$, para todo $v \in \mathbb{R}^n$ ■

Lema 3.23. Con las definiciones anteriores

$$\lim_{\varepsilon \downarrow 0} q_\varepsilon = q_0$$

Prueba.- Sea $\delta > 0$ y $v \in \mathbb{R}^n$ dados. Como cada h_i es localmente Lipschitz en x y la función $h_i^{\circ}(\cdot\,;\cdot)$ es continua superiormente, podemos escoger un $\varepsilon > 0$ de tal forma que cada h_i es Lipschitz en $B(x,\varepsilon)$ con la misma constante K, y que para todo $i = 1, \ldots, m$ tenemos

$$h_i^{\circ}(x_i; \pm v) \leq h_i^{\circ}(x; \pm v) + \delta/K \quad \text{para todo} \quad x_i \in B(x,\varepsilon)$$

Si $\xi_i \in \partial h_i(x_i)$ entonces $|\xi_i| \leq K$ para todo $i = 1, \ldots, m$. También se sabe que $h_i^{\circ}(y; \cdot)$ es positivamente homogénea, así que si multiplicamos por $|\xi_i|$ tenemos

$$h_i^{\circ}(x_i; \xi_i v) \leq h_i^{\circ}(x; \xi_i v) + |\xi_i|\delta/K \leq h_i^{\circ}(x; \xi_i v) + \delta$$

Por otro lado, sabiendo que el mapeo $\partial g(\cdot)$ es semicontinuo superiormente, podemos escoger el ε lo suficientemente pequeño como para que $\partial g(B(h(x);\varepsilon)) \subset B(\partial g(h(x)); \delta)$. Podemos calcular entonces

$$\begin{aligned}
q_0 &\leq q_\varepsilon(v)\\
&= \max\Big\{\sum_{i=1}^{m}\alpha_i\langle\xi_i,v\rangle \mid \xi_i\in\partial h_i(x_i),\alpha\in\partial g(u),x_i\in B(x,\varepsilon),u\in B(h(x),\varepsilon)\Big\}\\
&\leq \max\Big\{\sum_{i=1}^{m}\max\{\alpha_i\langle\xi_i,v\rangle \mid \xi_i\in\partial h_i(x_i),x_i\in B(x;\varepsilon)\} \mid \alpha\in B(\partial g(h(x));\delta)\Big\}\\
&\leq \max\Big\{\sum_{i=1}^{m}(h_i^\circ(x;\alpha_i v)+\delta) \mid \alpha\in B(\partial g(h(x));\delta)\Big\}\\
&\leq \max\Big\{\sum_{i=1}^{m}\max\{\alpha_i\langle\xi_i,v\rangle \mid \xi_i\in\partial h_i(x)\} \mid \alpha\in B(\partial g(h(x));\delta)\Big\}-m\delta\\
&\leq q_0+m\delta K|v|+m\delta\to q_0 \quad \text{cuando} \quad \delta\to 0
\end{aligned}$$

lo que completa la prueba. ■

Regresemos ahora a la prueba de la regla de la cadena. Vamos a demostrar que se cumple la desigualdad (3.20). Sea $\varepsilon>0$. Por definición de derivada direccional generalizada, existen $x'\in\mathbb{R}^n$ y $t>0$ tales que

$$f^\circ(x;v)\leq\frac{f(x'+tv)-f(x')}{t}+\varepsilon \tag{3.21}$$

y

$$\begin{cases} x',x'+tv\in B(x;\varepsilon)\\ h(x'),h(x'+tv)\in B(h(x);\varepsilon)\end{cases}$$

Por el teorema del valor medio, existe un $\alpha\in\partial g(u)$ tal que $u\in[h(x'+tv),h(x')]\subset B(h(x);\varepsilon)$ y

$$\begin{aligned}
f(x'+tv)-f(x') &= g(h(x'+tv))-g(h(x'))=\langle\alpha,h(x'+tv)-h(x')\rangle\\
&= \sum_{i=1}^{m}\alpha_i[h_i(x'+tv)-h_i(x')]
\end{aligned}$$

y aplicando nuevamente el teorema del valor medio para las funciones h_i, $i=1,\ldots,m$. Entonces existen subgradientes $\xi_i\in\partial h_i(x_i)$ tales que $\xi_i\in[x'+tv,x']\subset B(x;\varepsilon)$ y

$$\begin{aligned}
f(x'+tv)-f(x') &= \sum_{i=1}^{m}\alpha_i[h_i(x'+tv)-h_i(x')]=\sum_{i=1}^{m}\alpha_i\langle\xi_i,x'+tv-x'\rangle\\
&= \sum_{i=1}^{m}\alpha_i\langle\xi_i,tv\rangle
\end{aligned}$$

Ahora, de (3.21) se sigue que

$$\begin{aligned} f^{\circ}(x;v) &\leq \frac{\sum_{i=1}^{m}\alpha_i\langle\xi_i,tv\rangle}{t}+\varepsilon=\sum_{i=1}^{m}\alpha_i\langle\xi_i,tv\rangle+\varepsilon \\ &\leq q_{\varepsilon}(v)+\varepsilon \quad \text{para todo} \quad v\in\mathbb{R}^n \end{aligned} \tag{3.22}$$

lo que establece (3.20) y la afirmación básica.

Vamos a mostrar (i) ahora. Supongamos que g es regular en $h(x)$, h_i es regular en x y $\alpha_i > 0$ para todo $i = 1,\ldots,m$. Para probar la otra inclusion en (3.18) es suficiente con mostrar que

$$f^{\circ}(x;v)=q_0(v) \quad \text{para todo} \quad v\in\mathbb{R}^n$$

Ya sabemos que $f^{\circ}(x;v)\leq q_0(v)$ para todo $v\in\mathbb{R}^n$. Por el otro lado, tiendo que $\alpha_i\geq 0$ para todo $i=1,\ldots,m$, que cada h_i es regular en x y que g es regular en $h(x)$, implica que

$$\begin{aligned} q_0(v) &= \text{máx}\Big\{\sum_{i=1}^{m}\alpha_i\langle\xi_i,v\rangle \mid \xi_i\in\partial h_i(x),\alpha\in\partial g(h(x))\Big\} \\ &\leq \text{máx}\Big\{\sum_{i=1}^{m}\alpha_i\max_{\xi_i\in\partial h_i(x)}\langle\xi_i,v\rangle \mid \alpha\in\partial g(h(x))\Big\} \\ &= \text{máx}\Big\{\sum_{i=1}^{m}\alpha_i h_i^{\circ}(x;v) \mid \alpha\in\partial g(h(x))\Big\} \\ &= \text{máx}\Big\{\sum_{i=1}^{m}\alpha_i h_i'(x;v) \mid \alpha\in\partial g(h(x))\Big\} \\ &= g^{\circ}(h(x);h'(x;v))=g'(h(x);w), \end{aligned}$$

donde $w_i=h_i'(x;v)$. Entonces, por la definición de derivada direccional

$$\begin{aligned} g'(h(x);w) &= \lim_{t\downarrow 0}\frac{g(h(x)+tw)-g(h(x))}{t} \\ &= \lim_{t\downarrow 0}\Big\{\frac{g(h(x)+tv)-g(h(x))}{t}+T\Big\} \end{aligned}$$

donde $T=\dfrac{g(h(x)+tw)-g(h(x+tv))}{t}$. Vamos a obtener una cota superior de T y mostrar que ella tiende a cero cuando $t\to 0$. Usando la condición de Lipschitz de la función g tenemos

$$\begin{aligned} T &\leq \frac{|g(h(x)+tw)-g(h(x+tv))|}{t}\leq\frac{K\|h(x)+tw-h(x+tv)\|}{t} \\ &= K\Big\|w-\frac{h(x+tv)-h(x)}{t}\Big\|\longrightarrow K\|h'(x;v)-h'(x;v)\|=0, \quad \text{cuando} \quad t\to 0 \end{aligned}$$

Por esto,

$$q_0(v)\leq\lim_{t\downarrow 0}\frac{g(h(x+tv))-g(h(x))}{t}=f'(x;v)\leq f^{\circ}(x;v)$$

y por (3.22) tenemos que $q_0(v) = f'(x;v) = f^\circ(x;v)$ para todo $v \in \mathbb{R}^n$. Lo que quiere decir que f es regular y que la igualdad en (3.18) se cumple. Para mostrar (ii), supongamos que g es regular en $h(x)$ y que cada h_i es continuamente diferenciable en x para todo $i = 1, \dots, m$. Entonces, sabiendo que si una función es continuamente diferenciable, su subdiferencial es igual a su gradiente, tenemos que

$$\begin{aligned}
q_0(v) &= \text{máx}\Big\{ \sum_{i=1}^{m} \alpha_i \langle \xi_i, v\rangle \mid \xi_i \in \partial h_i(x) = \{\nabla h_i(x)\},\ \alpha \in \partial g(h(x))\Big\} \\
&= \text{máx}\Big\{ \sum_{i=1}^{m} \alpha_i \langle \nabla h_i(x), v\rangle \mid \alpha \in \partial g(h(x))\Big\} \\
&= \text{máx}\Big\{ \sum_{i=1}^{m} \alpha_i h_i'(x, v) \mid \alpha \in \partial g(h(x))\Big\}
\end{aligned}$$

y a partir de acá continuamos como en la demostración de (i) para obtener que $q_0(v) \le f^\circ(x;v)$, demostrando así la afirmación (ii).

Finalmente, para demostrar (iii), supongamos que $m = 1$ y que la función g es continuamente diferenciable en $h(x)$. Entonces

$$\alpha = g'(h(x)) = \lim_{y\to x} \frac{g(h(x)) - g(h(y))}{h(x) - h(y)}$$

y $\text{lím}_{x'\to x} g'(h(x')) = \alpha$. Podemos asumir, sin pérdida de generalidad, que $\alpha \ge 0$. Entonces calculamos

$$\begin{aligned}
q_0(v) &= \text{máx}\{\alpha\langle \xi, v\rangle \mid \xi \in \partial h(x)\} = \alpha \cdot h^\circ(x;v) \\
&= \limsup_{\substack{x'\to x\\ t\downarrow 0}} \frac{\alpha[h(x'+tv) - h(x')]}{t} = \limsup_{\substack{x'\to x\\ t\downarrow 0}} \frac{g'(h(x'))[h(x'+tv) - h(x')]}{t} \\
&= \limsup_{\substack{x'\to x\\ t\downarrow 0}} \frac{g(h(x'+tv)) - g(h(x'))}{t} = f^\circ(x;v) \quad \text{para todo} \quad v \in \mathbb{R}^n
\end{aligned}$$

lo que demuestra (iii) y así queda demostrado el teorema. ∎

Teorema 3.24 (Productos). Sean f_1 y f_2 localmente Lipschitz en $x \in \mathbb{R}^n$. Entonces la función $f_1 f_2$ es localmente Lipschitz en x y

$$\partial(f_1 f_2)(x) \subset f_2(x)\partial f_1(x) + f_1(x)\partial f_2(x) \tag{3.23}$$

Si además $f_1(x), f_2(x) \ge 0$ y f_1, f_2 son ambas regulares en x, entonces la igualdad se da y la función $f_1 f_2$ es regular en x.

<u>*Prueba*</u>.- Definamos la función $g : \mathbb{R}^2 \to \mathbb{R}$ por $g(u_1, u_2) = u_1 \cdot u_2$. Entonces g es continuamente diferenciable, entonces localmente Lipschitz en $(f_1(x), f_2(x))$ y $\partial f(x) = \{\nabla f(x)\}$ con

$$\partial g(f_1(x), f_2(x)) = \{\nabla g(f_1(x), f_2(x))\} = \{(f_2(x), f_1(x))\}$$

Ahora definamos la función $h : \mathbb{R}^n \to \mathbb{R}^2$ por $h(x) = (f_1(x), f_2(x))$. Ahora tenemos que $f_1 \cdot f_2 = g \circ h$. Por la regla de la cadena, la función $f_1 \cdot f_2$ es localmente Lipschitz en x y

$$\begin{aligned} \partial(f_1 f_2)(x) &\subset \operatorname{conv}\left\{ \sum_{i=1}^{2} \alpha_i \xi_i \mid \xi_i \in \partial h_i(x),\, \alpha \in \partial g(h(x)) \right\} \\ &= \operatorname{conv}\{f_2(x)\partial f_1(x) + f_1(x)\partial f_2(x)\} \end{aligned}$$

Entonces, como el $\partial f(x)$ es un conjunto convexo, el conjunto $f_2(x)\partial f_1(x) + f_1(x)\partial f_2(x)$ es convexo, entonces

$$\partial(f_1 f_2)(x) \subset f_2(x)\partial f_1(x) + f_1(x)\partial f_2(x)$$

Supongamos ahora que $f_1(x), f_2(x) \geq 0$ y que f_1, f_2 son regulares en x. Como la función g es continuamente diferenciable, es regular. Entonces, por el mismo teorema de la regla de la cadena, $f_1 f_2$ es regular, y la igualdad en (3.23) se da. ∎

Teorema 3.25 (Cocientes). Sea f_1 y f_2 localmente Lipschitz en $x \in \mathbb{R}^n$ y $f_2(x) \neq 0$. Entonces la función f_1/f_2 es localmente Lipschitz en x y

$$\partial\Big(\frac{f_1}{f_2}\Big)(x) \subset \frac{f_2(x)\partial f_1(x) - f_1(x)\partial f_2(x)}{f_2^2(x)} \tag{3.24}$$

Si además $f_1(x) \geq 0, f_2(x) \geq 0$ y f_1, f_2 son regulares en x, entonces la igualdad se da y la función f_1/f_2 es regular en x.

<u>*Prueba*</u>.- Definamos la función $g : \mathbb{R} \times (\mathbb{R} - \{0\}) \to \mathbb{R}$ por $g(u_1, u_2) = u_1/u_2$. Esta función es continuamente diferenciable, por lo que es localmente Lipschitz. Entonces, en $(f_1(x), f_2(x))$

$$\partial g(f_1(x), f_2(x)) = \{\nabla g(f_1(x), f_2(x))\} = \{(\frac{1}{f_2(x)}, -\frac{f_1(x)}{f_2^2(x)})\}$$

Si definimos $h : \mathbb{R}^n \to \mathbb{R}^2$ por $h(x) = (f_1(x), f_2(x))$, tenemos que $f_1/f_2 = g \circ h$. Usamos nuevamente la regla de la cadena, y tenemos que

$$\begin{aligned} \partial(f_1/f_2)(x) &\subset \operatorname{conv}\left\{ \sum_{i=1}^{2} \alpha_i \xi_i \mid \xi_i \in \partial h_i(x),\, \alpha \in \partial g(h(x)) \right\} \\ &= \operatorname{conv}\{\frac{1}{f_2(x)}\partial f_1(x) - \frac{f_1(x)}{f_2^2(x)}\partial f_2(x)\} \\ &= \operatorname{conv}\{\frac{f_2(x)\partial f_1(x) - f_1(x)\partial f_2(x)}{f_2^2(x)}\} \end{aligned}$$

El resto de la demostración es análogo al de productos. ∎

El siguiente teorema trata con una clase de funciones que son frecuentemente encontradas en optimización no suave, llamadas *funciones máximo*. El problema de minimizar ese tipo de función es usualmente llamado *problema de mínimo máximo*.

Teorema 3.26. Sea $f_i : \mathbb{R}^n \to \mathbb{R}$ localmente Lipschitz en x para cada $i = 1, \ldots, m$. Entonces la función $f : \mathbb{R}^n \to \mathbb{R}$ definida por

$$f(x) = \text{máx}\{f_i(x) \mid i = 1, \ldots, m\}$$

es localmente Lipschitz en x y

$$\partial f(x) \subset \text{conv}\,\{\partial f_i(x) \mid i \in I(x)\}, \tag{3.25}$$

donde $I(x) = \{i \in \{1, \ldots, m\} \mid f_i(x) = f(x)\}$. Si además, f_i es regular en x para cada i = 1,. . . ,m, entonces f es regular en x y la igualdad se da en (3.25).

<u>*Prueba*</u> .- Dada la definición de f, vemos que es de Lipschitz también. Definamos $g : \mathbb{R}^m \to \mathbb{R}$ y $h : \mathbb{R}^n \to \mathbb{R}^m$ por

$$\begin{aligned} g(u_1, \ldots, u_m) &= \underset{i=1,\ldots,m}{\text{máx}}\{u_i\} \\ h(x) &= (f_1(x), \ldots, f_m(x)) \end{aligned}$$

Tenemos entonces que $f = g \circ h$. Para todo $u, v \in \mathbb{R}^n$ y $\lambda \in [0, 1]$ se cumple

$$\begin{aligned} g(\lambda u + (1-\lambda)v) &= \underset{i=1,\ldots,m}{\text{máx}}\{\lambda u_i + (1-\lambda)v_i\} \\ &\leq \lambda \underset{i=1,\ldots,m}{\text{máx}}\{u_i\} + (1-\lambda) \underset{i=1,\ldots,m}{\text{máx}}\{v_i\} \\ &= \lambda g(u) + (1-\lambda)g(v), \end{aligned}$$

lo que significa que g es convexa y entonces, localmente Lipschitz en $h(x)$. Sea $J(u) = \{i \in \{1, \ldots, m\} \mid u_i = g(u)\}$. Entonces la derivada direccional es

$$\begin{aligned} g'(u; v) &= \lim_{t \downarrow 0} \frac{g(u+tv) - g(u)}{t} = \lim_{t \downarrow 0} \underset{i=1,\ldots,m}{\text{máx}} \frac{\{u_i + tv_i\} - g(u)}{t} \\ &= \lim_{t \downarrow 0} \underset{i \in J(u)}{\text{máx}} \frac{\{u_i + tv_i\} - g(u)}{t} \end{aligned}$$

Entonces

$$g'(u; v) = \underset{i \in J(u)}{\text{máx}}\, v_i$$

y como $g^{\circ} = g'$ (por ser convexa), tenemos que

$$\partial g(u) = \{\alpha \in \mathbb{R}^m \mid \max_{i \in J(u)} v_i \geq \langle \alpha, v \rangle \quad \text{para todo} \quad v \in \mathbb{R}^n\}$$

lo que quiere decir que

$$\alpha \in \partial g(u) \quad \Longleftrightarrow \quad \begin{cases} \alpha_i \geq 0, & i = 1, \ldots, m, \\ \sum_{i=1}^{m} \alpha_i = 1, & \\ \alpha_i = 0, & i \notin J(u). \end{cases}$$

entonces el subdiferencial de g en $h(x)$ es

$$\partial g(h(x)) = \{\alpha \in \mathbb{R}^n \mid \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i = 1 \quad \text{y} \quad \alpha_i = 0 \quad \text{si} \quad i \notin I(x)\}$$

Aplicando la regla de la cadena, tenemos

$$\begin{aligned} \partial f(x) &\subset \operatorname{conv} \Big\{ \sum_{i=1}^{m} \alpha_i \xi_i \mid \xi_i \in \partial h_i(x),\, \alpha \in \partial g(h(x)) \Big\} \\ &= \operatorname{conv} \Big\{ \sum_{i \in I(x)} \alpha_i \partial f_i(x) \mid \alpha_i \geq 0 \quad \text{y} \quad \sum_{i \in I(x)} \alpha_i = 1 \Big\} \\ &= \operatorname{conv} \{\partial f_i(x) \mid i \in I(x)\} \end{aligned}$$

Supongamos ahora que f_i es regular en x para todo $i \in I(x)$. Como g es convexa, es regular en $h(x)$. Entonces, la regla de la cadena implica la igualdad en (3.25). ■

Corolario 3.27. Supongamos que las funciones $f_i : \mathbb{R}^n \to \mathbb{R}$ son continuamente diferenciables en x y $g_i : \mathbb{R}^n \to \mathbb{R}$. Definamos las funciones $f : \mathbb{R}^n \to \mathbb{R}$ y $g : \mathbb{R}^n \to \mathbb{R}$ por

$$\begin{aligned} f(x) &= \max\{f_i(x) \mid i = 1, \ldots, m\} \quad \text{y} \\ g(x) &= \max\{g_i(x) \mid i = 1, \ldots, m\} \end{aligned}$$

Entonces tenemos

$$\partial f(x) = \operatorname{conv} \{\nabla f_i(x) \mid i = I(x)\} \quad \text{y} \tag{3.26}$$

$$\partial g(x) = \operatorname{conv} \{\partial_c g_i(x) \mid i = J(x)\} \tag{3.27}$$

donde $I(x) = \{i \in \{1, \ldots, m\} \mid f_i(x) = f(x)\}$ y $J(x) = \{i \in \{1, \ldots, m\} \mid g_i(x) = g(x)\}$

<u>Prueba</u>.- Se obtiene directamente usando el teorema anterior. ■

Teorema 3.28 (Rademacher). Sea $U \subset \mathbb{R}^n$ un conjunto abierto y $F : U \to \mathbb{R}$ Lipschitz de rango K en U. Entonces f es diferenciable en casi todas partes en U.

Prueba.- Usando la condición de Lipschitz, para todo $v \in \mathbb{R}^n$ tenemos

$$\limsup_{t\downarrow 0}\left|\frac{f(x+tv)-f(x)}{t}\right| \leq \limsup_{t\downarrow 0}\frac{tK\|v\|}{t} = K\|v\| < +\infty$$

lo que significa que f es diferenciable en casi todas partes. ■

Vamos a denotar por Ω_f al conjunto donde la función f es no diferenciable.

El siguiente Lema nos servirá para demostrar el teorema principal de esta sección.

Lema 3.29. Sea $\varepsilon > 0$ y $v \in \mathbb{R}^n$ tal que $v \neq 0$. Entonces

$$f^{\circ}(x;v) - \varepsilon \leq \limsup\{\langle \nabla f(y), v\rangle \mid y \to x,\ y \notin \Omega_f\} \tag{3.28}$$

Prueba.- Sea $\alpha = \limsup\{\langle \nabla f(y), v\rangle \mid y \to x,\ y \notin \Omega_f\}$. Entonces, por definición, existe $\delta > 0$ del forma que las condiciones

$$y \in B(x;\delta) \quad \text{y} \quad y \notin \Omega_f$$

implican $\langle \nabla f(y), v\rangle \leq \alpha+\varepsilon$. Tomemos este δ lo suficientemente pequeño como para que Ω_f tenga medida cero en $B(x;\delta)$. Ahora consideremos el segmento de línea $L_y = \{y+tv \mid 0 < t < \delta/(2|v|)\}$. Como Ω_f tiene medida cero en $B(x;d)$, se sigue del teorema de Fubini que para casi todo $y \in B(x;\delta/2)$, el segmento de línea L_y interseca a Ω_f en un conjunto de medida cero 1-dimensional. Sea y cualquier punto en $B(x;\delta/2)$ con esta propiedad y sea t dentro de $\langle 0, \delta/2|v|\rangle$. Entonces, por la definición de ∇ tenemos

$$f(y+tv) - f(y) = \int_0^t \langle \nabla f(y+sv), v\rangle ds,$$

ya que ∇f existe en casi todos los lados en L_y. Como tenemos que $|y+sv-x| < \delta$ para $0 < s < t$, de la definición inicial se tiene que $\langle \nabla f(y+sv), v\rangle \leq \alpha+\varepsilon$, así que

$$f(y+tv) - f(y) \leq t(\alpha+\varepsilon)$$

Como esto es cierto para cualquier y que se encuentre a una distancia no mayor de $\delta/2$ respecto a x y para todo t en $\langle 0, \delta/(2|v|)\rangle$, y como f es continua, en particular tenemos que

$$f^{\circ}(x;v) \leq \alpha + \varepsilon$$

■

Ahora el teorema principal:

Teorema 3.30. Sea $f : \mathbb{R}^n \to \mathbb{R}$ localmente Lipschitz en $x \in \mathbb{R}^n$. Entonces

$$\partial f(x) = \text{conv}\,\{\xi \in \mathbb{R}^n \mid \nabla f(x_i) \to \xi,\, x_i \to x \quad \text{y} \quad x_i \notin \Omega_f\} \tag{3.29}$$

<u>*Prueba*</u> .- Sea

$$A = \{\xi \in \mathbb{R}^n \mid \nabla f(x_i) \to \xi; x_i \to x \quad \text{y} \quad x_i \notin \Omega_f\}$$

Vamos a demostrar primero que A es no vacío. Se sigue del teorema de Rademacher que la medida del conjunto Ω_f es zero. Entonces existe una sucesión $\{x_i\} \subset \mathbb{R}^n$ tal que $x_i \notin \Omega_f$ y $x_i \to x$. Como f es localmente Lipschitz en x, existe $\varepsilon > 0$ tal que para todo $x_i \in B(x;\varepsilon)$

$$\|\xi_i\| \leq K \quad \text{para todo} \quad \xi_i \in \partial f(x_i) \tag{3.30}$$

Esto quiere decir que el mapeo punto-conjunto ∂f es localmente acotado en $B(x;\varepsilon)$. Sabemos que $\nabla f(x_i) \in \partial f(x_i)$, entonces por (3.30), la sucesión $\{\nabla f(x_i)\}$ es acotada. Entonces, por el Teorema de Bolzano-Weierstrass, la sucesión $\{\nabla f(x_i)\}$ admite una subsucesión convergente $\{\nabla f(x_{i_k})\}$, por lo que existe $\xi \in \mathbb{R}^n$ tal que $\nabla f(x_{i_k}) \to \xi$. Como ∂f es semicontinua por arriba, entonces $\xi \in \partial f(x)$. Hemos probado que A es no vacío, acotado y $A \subset \partial f(x)$. Como $\partial f(x)$ es convexo, entonces tenemos

$$\text{conv}\ A \subset \text{conv}\ \partial f(x) = \partial f(x)$$

Para la otra inclusión mostraremos que A también es cerrado, y así compacto. Sea $\{\xi_j\} \in A$ un sucesión tal que $\xi_j \to \xi$. Entonces

$$\xi_j = \lim_{i\to\infty} \nabla f(x_i^j), \quad \text{donde} \quad x_i^j \to x \quad \text{cuando} \quad i \to \infty \quad \text{y} \quad x_i^j \notin \Omega_f$$

Extrayendo subsucesiones si es necesario, existen puntos $x_i \in \mathbb{R}^n$ tales que

$$x_i = \lim_{j\to\infty} x_i^j \quad \text{para cada} \quad i \in \mathbb{N}$$

Entonces tenemos que $x_i \to x$, $x_i \notin \Omega_f$ y

$$\xi = \lim_{j\to\infty}\lim_{i\to\infty} \nabla f(x_i^j) = \lim_{i\to\infty}\lim_{j\to\infty} \nabla f(x_i^j) = \lim_{i\to\infty} \nabla f(x_i)$$

Por lo que $\xi \in A$, siendo así A cerrado y por lo tanto, compacto. Esto implica que su cápsula convexa conv A también es compacta , y usando el lema 3.29 y el teorema 3.15 concluimos con la prueba. ∎

3.4. ε-subdiferenciales

En optimización no suave, los llamados métodos paquete están basados en la ε-subdiferencial, que es la generalización del subdiferencial. Vamos a definirla y a presentar algunas propiedades básicas.

3.4.1. Caso Convexo

Vamos a generalizar la derivada direccional ordinaria.

Definición 3.31. Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa. La ***ε-derivada direccional de f*** en x en la dirección de $v \in \mathbb{R}^n$ esta definida por

$$f'_\varepsilon(x; v) = \inf_{t>0} \frac{f(x+tv) - f(x) + \varepsilon}{t} \tag{3.31}$$

Teorema 3.32. Sea f un función convexa. Entonces en cualquier punto $x \in \mathbb{R}^n$:

(i) la función $v \to f'_\varepsilon(x; v)$ es positivamente homogenea y subaditiva en $\mathbb{R}^n$ con

$$|f'_\varepsilon(x; v)| \leq K\|v\|$$

(ii) $f'_\varepsilon(x; v)$ es semicontinua por arriba como función de $(x; v)$ y Lipschitz con constante K como función de v en $\mathbb{R}^n$

(iii) $-f'_\varepsilon(x; -v) \leq f'_\varepsilon(x; v)$

Prueba .-

(i) Usando la condición de Lipschitz en $x \in \mathbb{R}^n$ y sabiendo que $\inf_{t>0} \frac{\varepsilon}{t} = 0$, tenemos

$$\begin{aligned} |f'_\varepsilon(x; v)| &\leq \inf_{t>0} \frac{|f(x+tv) - f(x) + \varepsilon|}{t} \leq \inf_{t>0} \frac{|f(x+tv) - f(x)| + \varepsilon}{t} \\ &\leq \inf_{t>0} \frac{K\|x+tv-x\| + \varepsilon}{t} = K\|v\| + \inf_{t>0} \frac{\varepsilon}{t} = K\|v\| \end{aligned}$$

Para ver que es positivamente homógenea, sea $\lambda > 0$. Entonces

$$\begin{aligned} f'_\varepsilon(x; \lambda v) &= \inf_{t>0} \frac{f(x+t\lambda v) - f(x) + \varepsilon}{t} = \inf_{t>0} \lambda \left\{ \frac{f(x+t\lambda v) - f(x) + \varepsilon}{\lambda t} \right\} \\ &= \lambda \inf_{t>0} \frac{f(x+t\lambda v) - f(x) + \varepsilon}{\lambda t} = \lambda f'_\varepsilon(x; v) \end{aligned}$$

Y ahora vamos a ver la subaditividad. Sean $v, w \in \mathbb{R}^n$ arbitrarios, entonces por la convxidad

$$\begin{aligned} f'_\varepsilon(x; v+w) &= \inf_{t>0} \frac{f(x+t(v+w)) - f(x) + \varepsilon}{t} \\ &= \inf_{t>0} \frac{f(\frac{1}{2}(x+2tv) + \frac{1}{2}(x+2tw)) - f(x) + \varepsilon}{t} \\ &\leq \inf_{t>0} \frac{f(x+2tv) - f(x) + \varepsilon}{2t} + \inf_{t>0} \frac{f(x+2tw) - f(x) + \varepsilon}{2t} \\ &= f'_\varepsilon(x; v) + f'_\varepsilon(x; w). \end{aligned}$$

(ii) Para ver la semicontinuidad por arriba, sean $\{x_i\}, \{v_i\} \subset \mathbb{R}^n$ sucesiones tales que $x_i \to x, v_i \to v$ y sea K la constante de Lipschitz en x. Definamos convenientemente la sucesión $\{t_i\} \subset \mathbb{R}$ por $t_i = K\|x_i - x\|^{\frac{1}{2}} + \frac{1}{i}$, por lo que $t_i \to 0$ y $t_i > 0$. Ahora tenemos

$$\begin{aligned} f'_\varepsilon(x_i; v_i) &= \inf_{t>0} \frac{f(x_i + tv_i) - f(x_i) + \varepsilon}{t} \leq \frac{f(x_i + t_i v_i) - f(x_i) + \varepsilon}{t_i} \\ &= \frac{f(x + t_i v) - f(x) + \varepsilon}{t_i} + \frac{f(x_i + t_i v_i) - f(x + t_i v)}{t_i} + \frac{f(x) - f(x_i)}{t_i} \end{aligned}$$

y si hacemos $i \to \infty$, tenemos por la condición de Lipschitz que

$$\begin{aligned} \frac{|f(x_i + t_i v_i) - f(x + t_i v)|}{t_i} &\leq \frac{K\|x_i - x\| + K\|t_i v_i - t_i v\|}{t_i} \\ &\leq \|x_i - x\|^{\frac{1}{2}} + K\|v_i - v\| \longrightarrow 0 \end{aligned}$$

y

$$\frac{|f(x) - f(x_i)|}{t_i} \leq \frac{K\|x - x_i\|}{t_i} \leq \|x - x_i\|^{\frac{1}{2}} \longrightarrow 0$$

Tomando límites por arriba (cuando $i \to \infty$),obtenemos

$$\limsup_{i\to\infty} f'_\varepsilon(x_i, v_i) \leq \limsup_{i\to\infty} \frac{f(x + t_i v) - f(x) + \varepsilon}{t_i} = f'_\varepsilon(x; v)$$

Para mostrar la condición de Lipschitz, sean $v, w \in \mathbb{R}^n$ dados. Si $x + tv$ y $x + tw$ se encuentran suficientemente cercanos a x, entonces

$$f(x + tv) - f(x + tw) \leq Kt\|v - w\|$$

sumando y restando el término $f(x) - \varepsilon$ y dividiendo entre t tenemos

$$\inf_{t>0} \frac{f(x + tv) - f(x) + \varepsilon}{t} \leq \inf_{t>0} \frac{f(x + tw) - f(x) + \varepsilon}{t} + K\|v - w\|$$

por consiguiente

$$f'_\varepsilon(x; v) - f'_\varepsilon(x; w) \leq K\|v - w\|$$

Invirtiendo papeles entre v y w tenemos

$$f'_\varepsilon(x;w) - f'_\varepsilon(x;v) \leq K\|v-w\|$$

por lo que

$$|f'_\varepsilon(x;v) - f'_\varepsilon(x;w)| \leq K\|v-w\|$$

(iii) Usemos el item (i)

$$\tfrac{1}{2}f'_\varepsilon(x;v) + \tfrac{1}{2}f'_\varepsilon(x;-v) \geq f'_\varepsilon(x;\tfrac{1}{2}v - \tfrac{1}{2}v) = 0$$

lo que implica que $-f'_\varepsilon(x;-v) \leq f'_\varepsilon(x;v)$ ∎

Ahora vamos a generalizar el subgradiente y el subdiferencial de una función convexa.

<u>Definición</u> 3.33. Sea $\varepsilon \geq 0$, entonces el *ε-subdiferencial de una función convexa* $f : \mathbb{R}^n \to \mathbb{R}$ en $x \in \mathbb{R}^n$ es el conjunto

$$\partial_\varepsilon f(x) = \{\xi \in \mathbb{R}^n \mid f(x') \geq f(x) + \langle \xi, x'-x\rangle - \varepsilon \quad \text{para todo} \quad x' \in \mathbb{R}^n\} \tag{3.32}$$

Cada elemento ξ de $\partial_\varepsilon f(x)$ es un *ε-subgradiente de la función convexa* f en x.

El siguiente teorema resume la propiedades básicas del ε-subdiferencial

Teorema 3.34. Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Entonces

(i) $\partial_0 f(x) = \partial_c f(x)$

(ii) Si $\varepsilon_1 \leq \varepsilon_2$, entonces $\partial_{\varepsilon_1} f(x) \subset \partial_{\varepsilon_2} f(x)$

(iii) $f'_\varepsilon(x;v) = \max\{\langle \xi, v\rangle \mid \xi \in \partial_\varepsilon f(x)\}$ para todo $v \in \mathbb{R}^n$

(iv) $\partial_\varepsilon f(x) = \{\xi \in \mathbb{R}^n \mid f'_\varepsilon(x;v) \geq \langle \xi, v\rangle \quad \text{para todo} \quad v \in \mathbb{R}^n\}$

(v) $\partial_\varepsilon f(x)$ es un conjunto no vacío, convexo y compacto tal que $\|\xi\| \leq K$ para todo $\xi \in \partial_\varepsilon f(x)$

(vi) El mapeo $\partial_\varepsilon f(\cdot) : \mathbb{R}^n \to \mathcal{P}(R^n)$ es semicontinuo por arriba.

<u>Prueba</u> .-

(i) y (ii) se obtienen directamente de la definición.

(iii) Por la definición de ε-subdiferencial, deducimos que para cada $v \in \mathbb{R}^n$

$$\varphi_\varepsilon(t) = \frac{f(x+tv) - f(x) + \varepsilon}{t} \geq \frac{\langle \xi, tv \rangle}{t} = \langle \xi, v \rangle \quad \text{para todo} \quad \xi \in \partial_\varepsilon f(x)$$

entonces,

$$f'_\varepsilon(x; v) \geq \text{máx}\{\langle \xi, v \rangle \mid \xi \in \partial_\varepsilon f(x)\}$$

Por otro lado, si existiera un $v_1 \in \mathbb{R}^n$ para el cual $f'_\varepsilon(x; v_1) > \text{máx}\{\langle \xi, v_1 \rangle \mid \xi \in \partial_\varepsilon f(x)\}$, existiría, por el Teorema de Hahn-Banach (Teorema 3.14), algún $\xi_1 \in \mathbb{R}^n$ tal que

$$f'_\varepsilon(x; \cdot) \geq \langle \xi_1, \cdot \rangle \quad \text{y} \quad f'_\varepsilon(x; v_1) = \langle \xi_1 . v_1 \rangle$$

Para cualquier $x' \in \mathbb{R}^n$ existen $v \in \mathbb{R}^n$ y $t > 0$ tales que $x' = x + tv$. Entonces

$$f(x') - f(x) + \varepsilon = t\varphi_\varepsilon(t) \geq t f'_\varepsilon(x; v) \geq t\langle \xi_1, v \rangle = \langle \xi_1, tv \rangle = \langle \xi_1, x' - x \rangle \quad \forall x' \in \mathbb{R}^n$$

lo que implica que $\xi_1 \in \partial_\varepsilon f(x)$ y

$$f'_\varepsilon(x; v_1) > \text{máx}\{\langle \xi, v_1 \rangle \mid \xi \in \partial_\varepsilon f(x)\} \geq \langle \xi_1, v_1 \rangle = f'_\varepsilon(x, v_1).$$

lo que es imposible, por lo que se cumple la igualdad.

(iv) Sea $K = \{\xi \in \mathbb{R}^n \mid f'_\varepsilon(x; v) \geq \langle \xi, v \rangle \quad \text{para todo} \quad v \in \mathbb{R}^n\}$ y sea $\xi \in K$ arbitrario. Entonces por la convexidad se sigue que, para todo $y \in \mathbb{R}^n$, se tiene

$$\begin{aligned}
\langle \xi, y \rangle &\leq f'_\varepsilon(x; y) \\
&= \inf_{t>0} \frac{f((1-t)x + t(x+y)) - f(x) + \varepsilon}{t} \\
&\leq \inf_{t>0} \frac{(1-t)f(x) + tf(x+y) - f(x) + \varepsilon}{t} \\
&\leq f(x+y) - f(x) + \varepsilon
\end{aligned}$$

con $t \leq 1$. Como $\varepsilon > 0$, si hacemos $y = x' - x$, tendriamos que $\xi \in \partial_\varepsilon f(x)$

Por otro lado, si $\xi \in \partial_\varepsilon f(x)$ entonces para todo $y \in \mathbb{R}^n$

$$f'_\varepsilon(x; y) = \inf_{t>0} \frac{f(x+ty) - f(x) + \varepsilon}{t} \geq \inf_{t>0} \frac{\langle \xi, x + ty - x \rangle}{t} = \langle \xi, y \rangle$$

Así que $\xi \in K$.

(v) Por el teorema 3.32(i), la función $f'_\varepsilon(x; \cdot) : \mathbb{R}^n \to \mathbb{R}$ es positivamente homogénea y subaditiva, entonces por el Teorema 3.14, existe un vector $\xi \in \mathbb{R}^n$ tal que $\langle \xi, v \rangle \leq f'_\varepsilon(x; v)$

para todo $v \in \mathbb{R}^n$, lo que implica, por (iv), que $\partial_\varepsilon f(x)$ es no vacío. Para ver la convexidad, sean $\xi_1, \xi_2 \in \partial_\varepsilon f(x)$ y $\lambda \in [0,1]$. Entonces por (iii)

$$\begin{aligned}\langle \lambda\xi_1 + (1-\lambda)\xi_2, v\rangle &= \lambda\langle \xi_1, v\rangle + (1-\lambda)\langle \xi_2, v\rangle \\ &\leq \lambda f'_\varepsilon(x;v) + (1-\lambda) f'_\varepsilon(x;v) \\ &\leq f'_\varepsilon(x;v),\end{aligned}$$

así que por (iv) $\lambda\xi_1 + (1-\lambda)\xi_2 \in \partial_\varepsilon f(x)$.

Ahora, por (iv) y por el teorema 3.32(i), tenemos que para cualquier $\xi \in \partial_\varepsilon f(x)$

$$\|\xi\|^2 = |\langle \xi, \xi\rangle| \leq |f'_\varepsilon(x;\xi)| \leq K\|\xi\|$$

lo que implica que $\partial_\varepsilon f(x)$ es acotado, así que para demostrar que es compacto, es suficiente con mostrar que es cerrado. Sea $\{\xi_i\} \subset \partial_\varepsilon f(x)$ una sucesión tal que $\xi_i \to \xi$. Entonces

$$\langle \xi, v\rangle = \langle \lim_{i\to\infty} \xi_i, v\rangle = \lim_{i\to\infty} \langle \xi_i, v\rangle \leq \lim_{i\to\infty} f'_\varepsilon(x;v) = f'_\varepsilon(x;v), \qquad \forall v \in \mathbb{R}^n$$

lo que muestra que $\xi \in \partial_\varepsilon f(x)$, y por lo tanto $\partial_\varepsilon f(x)$ es cerrado.

(vi) Finalmente, sean $\{y_i\} \subset \mathbb{R}^n$ y $\{\xi_i\} \subset \partial_\varepsilon f(y_i)$ sucesiones tales que $y_i \to x$ y $\xi_i \to \xi$. Entonces para todo $v \in \mathbb{R}^n$, por (iv), tenemos que

$$\langle \xi, v\rangle = \langle \lim_{i\to\infty} \xi_i, v\rangle = \lim_{i\to\infty} \langle \xi_i, v\rangle \leq \limsup_{i\to\infty} \langle \xi_i, v\rangle \leq \limsup_{i\to\infty} f'_\varepsilon(y_i;v)$$

Por el teorema 3.32(ii), la función $f'_\varepsilon(x;\cdot)$ es semicontinua por arriba, por lo que

$$\langle \xi, v\rangle \leq f'_\varepsilon(x;v)$$

Y así, la función $\partial_\varepsilon f(\cdot)$ también es semicontinua por arriba.

■

Teorema 3.35. Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa con constante de Lipschitz K en x. Si $\varepsilon \geq 0$, entonces

$$\partial_c f(y) \subset \partial_\varepsilon f(x) \quad \text{para todo} \quad y \in B(x; \frac{\varepsilon}{2K}) \tag{3.33}$$

<u>*Prueba*</u>.- Sean $\xi \in \partial_c f(y)$ y $y \in B(x; \frac{\varepsilon}{2K})$. Entonces para todo $z \in \mathbb{R}^n$ se cumple

$$\begin{aligned} f(z) &\geq f(y) + \langle \xi, z-y\rangle \\ &= f(x) + \langle \xi, z-x\rangle - (f(x) - f(y) + \langle \xi, z-x\rangle - \langle \xi, z-y\rangle)\end{aligned}$$

usando la condición de Lipschitz y sabiendo que se cumplen análogamente los mismos enunciados del teorema 3.34 para $\partial_c f(x)$, tenemos que

$$\begin{aligned} |f(x) - f(y) + \langle \xi, z - x\rangle - \langle \xi, z - y\rangle| &\leq |f(x) - f(y)| + |\langle \xi, z - x\rangle - \langle \xi, z - y\rangle| \\ &\leq K\|x - y\| + \|\xi\|\|x - y\| \\ &\leq 2K\|x - y\| \leq 2K \cdot \frac{\varepsilon}{2K} = \varepsilon \end{aligned}$$

Reemplazando en lo anterior, obtenemos que $\xi \in \partial_\varepsilon f(x)$. ■

Este teorema lo que muestra es que el ε-subdiferencial contiene de forma condensada la información subgradiente de una vecindad.

3.4.2. Caso No Convexo

Regresemos al caso no convexo. Es posible generalizar la ε-subdiferencial análogamente para el caso de funciones de Lipschitz usando la derivada direccional generalizada. Sin embargo, es mas útil usar la ε-subdiferencial de *Goldstein* para funciones no convexas.

Definición 3.36. Sea $\varepsilon \geq 0$, entonces la ***ε-subdiferencial de Goldstein*** de una función Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ en $x \in \mathbb{R}^n$ es el conjunto

$$\partial_\varepsilon^G f(x) = \operatorname{conv}\{\partial f(y) \mid y \in B(x;\varepsilon)\} \tag{3.34}$$

Cada elemento $\xi \in \partial_\varepsilon^G f(x)$ es llamado un ε-subgradiente de la función f en x.

Teorema 3.37. Sea $f : \mathbb{R}^n \to \mathbb{R}$ localmente Lipschitz de rango K en x. Entonces

(i) $\partial_0^G f(x) = \partial f(x)$
(ii) Si $\varepsilon_1 \leq \varepsilon_2$, entonces $\partial_{\varepsilon_1}^G f(x) \subset \partial_{\varepsilon_2}^G f(x)$
(iii) $\partial_\varepsilon^G f(x)$ es un conjunto no vacío, convexo y compacto tal que $\|\xi\| \leq K$ para todo $\xi \in \partial_\varepsilon^G f(x)$
(iv) El mapeo $\partial_\varepsilon^G f(\cdot) : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ es semicontinuo superiormente.

Prueba .-

(i) y (ii) se obtienen directamente de la definición.

(iii) La afirmación (i) implica que para todo $\varepsilon \geq 0$

$$\partial f(x) = \partial_0^G f(x) \subset \partial_\varepsilon^G f(x)$$

De forma análoga a la demostración del teorema 3.34(v), se obtiene que $\partial f(x)$ es no vacío, y así $\partial_\varepsilon^G f(x)$ también es no vacío. Debido a la definición de $\partial_\varepsilon^G f(x)$, evidentemente es convexo. La compacidad se obtiene de la propiedad de compacto de $\partial f(x)$. Vamos a demostrar la desigualdad. Sea $\xi \in \partial_\varepsilon^G f(x)$ arbitrario. Entonces $\xi = \sum_{i=1}^m \lambda_i \xi_i$, donde $\xi_i \in \partial f(y_i)$, $y_i \in B(x;\varepsilon)$, $\lambda_i \geq 0$ y $\sum_{i=1}^m \lambda_i = 1$. Así tenemos que

$$\|\xi\| = \|\sum_{i=1}^m \lambda_i \xi_i\| \leq \sum_{i=1}^m \lambda_i \|\xi_i\| \leq \sum_{i=1}^m \lambda_i \cdot K = K$$

ya que $\|\xi_i\| \leq K \quad$ para todo $\quad i$.

(iv) Se obtiene directamente ya que $\partial f(x)$ tiene la misma propiedad. ∎

Teorema 3.38. Sea $f : \mathbb{R}^n \to \mathbb{R}$ localmente Lipschitz en $x \in \mathbb{R}^n$. Entonces

$$\partial_\varepsilon^G f(x) = \operatorname{conv}\{\xi \in \mathbb{R}^n \mid \nabla f(y_i) \to \xi,\, y_i \to y,\, y_i \notin \Omega_f,\, y \in B(x:\varepsilon)\} \tag{3.35}$$

<u>*Prueba*</u>.- Usando el teorema 3.30 se obtiene directamente. ∎

Teorema 3.39. Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa con constante de Lipschitz K en x. Entonces para todo $\varepsilon \geq 0$ tenemos que

$$\partial_\varepsilon^G f(x) \subset \partial_{2K\varepsilon} f(x) \tag{3.36}$$

<u>*Prueba*</u>.- Sea $\xi \in \partial_\varepsilon^G f(x)$. Entonces $\xi = \sum_{i=1}^m \lambda_i \xi_i$, donde $\xi_i \in \partial f(y_i)$, $y_i \in B(x;\varepsilon)$, $\lambda_i \geq 0$ y $\sum_{i=1}^m \lambda_i = 1$. Entonces para todo $i = 1, \ldots, m$ tenemos

$$f(x') \geq f(y_i) + \langle \xi_i, x' - y_i \rangle \quad \text{para todo} \quad x' \in R^n$$

multiplicamos por λ_i ambos lados y sumamos sobre i para obtener

$$\begin{aligned} f(x') &= \sum_{i=1}^m \lambda_i f(x') \geq \sum_{i=1}^m \lambda_i f(y_i) + \sum_{i=1}^m \lambda_i \langle \xi_i, x' - y_i \rangle \quad \text{para todo} \quad x' \in \mathbb{R}^n \\ &= f(x) + \langle \xi, x' - x \rangle - (f(x) - \sum_{i=1}^m \lambda_i f(y_i) + \langle \xi, x' - x \rangle - \sum_{i=1}^m \lambda_i \langle \xi_i, x' - y_i \rangle) \end{aligned}$$

y usando la condición de Lipschitz y sabiendo que $\|\xi\| \leq K$, tenemos

$$
\begin{aligned}
&|f(x) - \sum_{i=1}^{m} \lambda_i f(y_i) + \langle \xi, x' - x \rangle - \sum_{i=1}^{m} \lambda_i \langle \xi_i, x' - y_i \rangle| \\
&\leq |\sum_{i=1}^{m} \lambda_i f(x) - \sum_{i=1}^{m} \lambda_i f(y_i)| + |\sum_{i=1}^{m} \lambda_i \langle \xi_i, x' - x \rangle - \sum_{i=1}^{m} \lambda_i \langle \xi_i, x' - y_i \rangle| \\
&\leq \sum_{i=1}^{m} \lambda_i |f(x) - f(y_i)| + \sum_{i=1}^{m} \lambda_i |\langle \xi_i, x - y_i \rangle| \\
&\leq \sum_{i=1}^{m} \lambda_i (K\|x - y_i\| + \|\xi\| \|x - y_i\|) = 2K\varepsilon
\end{aligned}
$$

lo que significa, por defincíón, que $\xi \in \partial_{2K\varepsilon} f(x)$ ∎

Este teorema muestra la relación entre la ε-subdiferencial para funciones convexas y la ε-subdiferencial de Goldstein.

3.5. Jacobianos Generalizados

Supongamos que tenemos la función $F : \mathbb{R}^n \to \mathbb{R}^m$ escrita en términos de sus funciones componentes como $F(x) = (F_1(x), \ldots, F_m(x))^T$, en la que cada una de las F_i para $i = 1, \ldots, m$ son localmente Lipschitz. Así F también sería localmente Lipschitz y por el teorema 3.28 concluimos que es diferenciable en casi todas partes. Si denotamos por Ω_F al conjunto en $\mathbb{R}^n$ donde F no es diferenciable y $\nabla F(x)$ a la matriz Jacobiana usual $m \times n$ en $x \notin \Omega_F$, entonces, basados en el teorema 3.30 podemos generalizar la derivada de F.

Definición 3.40. Sea $F : \mathbb{R}^n \to \mathbb{R}^m$ localmente Lipschitz en un punto $x \in \mathbb{R}^n$. Entonces el *Jacobiano generalizado de F en x* es el conjunto

$$\partial F(x) = \operatorname{conv} \{Z \in \mathbb{R}^{m \times n} \mid \nabla F(x_i) \to Z, x_i \to x, \quad \text{y} \quad x_i \notin \Omega_F\} \tag{3.37}$$

Sabemos que el espacio de las matrices $m \times n$ está dotado con la norma

$$\|Z\|_{m \times n} = \Big(\sum_{i=1}^{m} \|Z_i\|^2 \Big)^{\frac{1}{2}} \tag{3.38}$$

donde Z_i es la i-ésima fila de Z. Con esta norma, podemos mencionar algunas propiedades básicas de $\partial F(x)$.

Corolario 3.41. Sean $F_i : \mathbb{R}^n \to \mathbb{R}$ localmente Lipschitz con constante K_i para $i = 1, \ldots, m$. Entonces:

(i) $F(x) = (F_1(x), \ldots, F_m(x))^T$ es localmente Lipschitz en x con constante $K = \|(K_1, \ldots, K_m)^T\|$

(ii) $\partial F(x)$ es un subconjunto no vacío, convexo y compacto de $\mathbb{R}^{m\times n}$

(iii) El mapeo $\partial F(\cdot) : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ es semicontinuo superiormente.

Prueba.- Directamente del teorema 3.30 y sabiendo que el teorema 3.34 también se cumple para funciones localmente Lipschitz. ■

Corolario 3.42. [Regla de la Cadena del Jacobiano] Sea $f = g \circ F$, donde $F : \mathbb{R}^n \to \mathbb{R}^m$ es localmente Lipschitz en $x \in \mathbb{R}^n$ y $g : \mathbb{R}^m \to \mathbb{R}$ es localmente Lipschitz en $F(x) \in \mathbb{R}^m$. Entonces f es localmente Lipschitz en x y

$$\partial f(x) \subset \operatorname{conv}\left\{\partial F(x)^T \partial g(F(x))\right\} \tag{3.39}$$

Si además g es continuamente diferenciable, entonces $\partial g(F(x)) = \{\nabla g(F(x))\}$ y

$$\partial f(x) = \partial F(x)^T \partial g(F(x)) \tag{3.40}$$

Prueba.- De la regla de la cadena 3.22 y el teorema 3.30 ■

3.6. Geometría No Suave

En esta sección vamos a demostrar que los conceptos geométricos pueden ser generalizados para el análisis no convexo y que hay una conexión entre ellos.

3.6.1. Generalización de Tangentes y Normales

Definición 3.43. Sea G un conjunto no vacío de $\mathbb{R}^n$. *El Cono Tangente* del conjunto G en $x \in G$ es dado por

$$T_G(x) = \{y \in \mathbb{R}^n \mid d_G^\circ(x; y) = 0\} \tag{3.41}$$

Este cono tangente tiene las mismas propiedades elementales que en el caso convexo.

Teorema 3.44. El Cono Tangente $T_G(x)$ de un conjunto G es un cono cerrado y convexo que contiene el cero.

Prueba.- Dado que el teorema 3.32 se cumple también para la derivada direccional generalizada, entonces podemos usar los items (i) y (ii). El item (i), implica que $T_G(x)$ es un cono.

Para la convexidad, sean $y_1, y_2 \in T_G(x)$ y $\lambda \in [0,1]$, entonces

$$\begin{aligned} d_G^{\circ}(x; \lambda y_1 + (1-\lambda)y_2) &\leq d_G^{\circ}(x; \lambda y_1) + d_G^{\circ}(x; (1-\lambda)y_2) \\ &= \lambda d_G^{\circ}(x; y_1) + (1-\lambda) d_G^{\circ}(x; y_2) = 0 + 0 = 0 \end{aligned}$$

y como de la definición de $d_G(x)$, $d_G{}^{\circ}(x;y)$ es no negativo, se tiene que $d_G^{\circ}(x; \lambda y_1 + (1-\lambda)y_2) = 0$ lo que demuestra la convexidad.

Para demostrar que es cerrado, sea $\{y_i\} \subset T_G(x)$ una sucesión que converge a $y \in \mathbb{R}^n$. Entonces,

$$\begin{aligned} 0 = \lim_{i\to\infty} d_G^{\circ}(x; y_i) &= \lim_{i\to\infty} \left\{ \limsup_{\substack{y' \to x \\ t \downarrow 0}} \frac{d_G(y' + t y_i) - d_G(y')}{t} \right\} = \lim_{i\to\infty} \left\{ \limsup_{\substack{y' \to x \\ t \downarrow 0}} \frac{\inf_{g\in G} \|y' + t y_i - g\| - d_G(y')}{t} \right\} \\ &= \limsup_{\substack{y' \to x \\ t \downarrow 0}} \frac{\inf_{g\in G} \|y' + t y - g\| - d_G(y')}{t} = d_G{}^{\circ}(x; y) \end{aligned}$$

Y para demostrar que contiene al cero, sea $y \in \mathbb{R}^n$, entonces

$$d_G^{\circ}(x; 0) \leq d_G^{\circ}(x; y) + d_G^{\circ}(x; -y) = d_G^{\circ}(x; y) + (-d_G^{\circ})(x; y) = d_G^{\circ}(x; y) - d_G^{\circ}(x; y) = 0$$

■

Como en el caso no convexo, usaremos la polaridad para definir el cono normal.

<u>**Definición**</u> **3.45.** Sea G un conjunto no vacío de $\mathbb{R}^n$. *El Cono Normal* del conjunto G en el punto $x \in G$ es el conjunto

$$N_G(x) = T_G(x)^{\circ} = \{ z \in \mathbb{R}^n \mid \langle y, z \rangle \leq 0 \quad \text{para todo} \quad y \in T_G(x) \} \tag{3.42}$$

Teorema 3.46. El Cono Normal $N_G(x)$ de un conjunto G es un cono cerrado y convexo que contiene el cero.

<u>Prueba</u>.- De la misma definición se obtiene rápidamente. ■

El siguiente teorema muestra que efectivamente las definiciones que hemos dado son generalizaciones del caso convexo.

Teorema 3.47. Si el conjunto G es convexo, entonces el cono tangente y el cono normal de G definidos anteriormente, coinciden con los conos tangente y normal respectivamente definidos para conjuntos convexos.

Prueba.- Sea $x \in G$. Sabemos que para conjuntos convexos

$$T_G(x) = \{y \in \mathbb{R}^n \mid d'_G(x;y) = 0\}$$

Entonces, como el conjunto G es convexo, la función d_G es convexa, y al ser convexa, su derivada direccional y su derivada direccional generalizada coinciden, esto es

$$d'_G(x;y) = d^\circ_G(x;y) \quad \text{para todo} \quad y \in \mathbb{R}^n$$

lo que demuestra el teorema. ∎

Teorema 3.48. El cono tangente del conjunto G en x también puede ser escrito como

$$T_G(x) = \{y \in \mathbb{R}^n \mid \quad \text{para todo} \quad t_j \downarrow 0 \quad \text{y} \quad x_j \to x \quad \text{con} \quad x_j \in G \quad \text{existen} \quad y_j \to y \quad \text{con} \quad x_j + t_j y_j \in G\} \tag{3.43}$$

Prueba.- Sea

$$S = \{y \in \mathbb{R}^n \mid \quad \text{para todo} \quad t_j \downarrow 0 \quad \text{y} \quad x_j \to x \quad \text{con} \quad x_j \in G \quad \text{existen} \quad y_j \to y \quad \text{con} \quad x_j + t_j y_j \in G\}$$

Supongamos que $y \in T_G(x)$, y que las sucesiones $x_j \to x$ con $x_j \in G$ y $t_j \downarrow 0$ están dadas. Como $d^\circ_G(x;y) = 0$ y $x_j \in G$, tenemos

$$\begin{aligned} 0 &\le \lim_{j\to\infty} \frac{d_G(x_j + t_j y)}{t_j} - \lim_{j\to\infty} \frac{d_G(x_j + t_j y) - d_G(x_j)}{t_j} \\ &\le \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{d_G(x' + ty) - d_G(x')}{t} = d^\circ_G(x;y) = 0 \end{aligned}$$

Entonces el límite existe y es cero. Entonces para todo $j \in \mathbb{N}$ existe un $g_j \in G$ tal que

$$\|x_j + t_j y - g_j\| \le d_G(x_j + t_j y) + \frac{t_j}{j}$$

y si definimos

$$y_j = \frac{g_j - x_j}{t_j}$$

Entonces,

$$\begin{aligned} \|y - y_j\| &= \left\| y - \frac{g_j - x_j}{t_j} \right\| \\ &= \frac{\|x_j + t_j y - g_j\|}{t_j} \le \frac{d_G(x_j + t_j y)}{t_j} + \frac{1}{j} \longrightarrow 0 \end{aligned}$$

cuando $j \to \infty$ y

$$x_j + t_j y_j = x_j + t_j \left(\frac{g_j - x_j}{t_j} \right) = g_j \in G$$

entonces $y \in S$.

Para la otra inclusión, supongamos que $y \in S$ y escojamos sucesiones $z_j \to x$ y $t_j \downarrow 0$ tales que

$$\lim_{j \to \infty} \frac{d_G(z_j + t_j y - d_G(z_j))}{t_j} = d_G^\circ(x; y) \tag{3.44}$$

Para probar que $d_G^\circ(x; y) = 0$, es suficiente con probar que la cantidad en (3.44) es no positiva. Para verlo, escojamos $g_j \in G$ tales que

$$\|g_j - z_j\| \leq d_G(z_j) + \frac{t_j}{j}$$

Entonces,

$$\|x - g_j\| \leq \|x - z_j\| + \|z_j - g_j\| \leq \|x - z_j\| + d_G(z_j) + \frac{t_j}{j} \longrightarrow 0$$

cuando $j \to \infty$. Hemos hallado un sucesión en G que converge a x, entonces por hipótesis existe un sucesión y_j que converge a y tal que $g_j + t_j y_j \in G$. Dado que la función d_G es de Lipschitz de rango 1, tenemos

$$\begin{aligned} d_G(z_j + t_j y) &\leq d_G(g_j + t_j y_j) + \|z_j - g_j\| + t_j \|y - y_j\| \\ &\leq d_G(z_j) + t_j \left(\|y - y_j\| + \frac{1}{j} \right) \end{aligned}$$

lo que implica que la cantidad en (3.44) es no positiva, completando así la prueba. ■

Teorema 3.49. El Cono Normal en x del conjunto G también puede ser escrito como

$$N_G(x) = \text{clausura} \left\{ \bigcup_{\lambda \geq 0} \lambda \, \partial d_G(x) \right\} \tag{3.45}$$

<u>*Prueba*</u>.- Sea $z \in \partial d_G(x)$ dado. Entonces, por definición de subdiferencial

$$\langle z, y \rangle \leq d_G^\circ(x; y) \quad \text{para todo} \quad y \in \mathbb{R}^n.$$

Si tenemos que $y \in T_G(x)$ entonces $d_G^\circ(x; y) = 0$. Entonces, $\langle z, y \rangle \leq 0$ para todo $y \in T_G(x)$ lo que implica que $z \in T_G(x)^\circ = N_G(x)$. Como $N_G(x)$ es un cono convexo y cerrado, tenemos

$$\text{clausura} \left\{ \bigcup_{\lambda \geq 0} \lambda \, \partial d_G(x) \right\} \subset N_G(x)$$

Para la otra inclusión, sea $z \in N_G(x)$. De las definiciones de conos tangente y normal, tenemos

$$\langle z, y \rangle \leq 0 = d_G^\circ(x; y) \quad \text{para todo} \quad y \in T_G(x).$$

Supongamos ahora que $y \notin T_G(x)$, entonces por el teorema 3.44, $y \neq 0$. Se sigue que

$$d_G^\circ(x;y) = \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{d_G(x'+ty) - d_G(x')}{t} \geq \limsup_{t \downarrow 0} \frac{1}{t} \cdot d_G(x+ty) \geq 0.$$

Si $z = 0$, entonces $\langle z, y \rangle = 0 \leq d_G^\circ(x;y)$, mientras que si $z \neq 0$ podemos escoger

$$\lambda_{z,y} = \frac{d_G^\circ(x;y)}{\|z\| \|y\|} \geq 0$$

tenemos entonces

$$\langle \lambda_{z,y} z, y \rangle \leq \lambda_{z,y} \|z\| \|y\| = d_G^\circ(x;y).$$

En consecuencia, para todo $z \in N_G(x)$ y $y \in \mathbb{R}^n$ tenemos un $\lambda_{z,y} \geq 0$ tal que (si $y \in T_G(x)$ o $z = 0$, $\lambda_{z,y} = 1$)

$$\langle \lambda_{z,y} z, y \rangle \leq d_G^\circ(x;y),$$

lo que significa que

$$z \in \bigcup_{\lambda \geq 0} \lambda \, \partial d_G(x)$$

lo que completa la prueba. ∎

Teorema 3.50. Si $x \in \operatorname{int} G$, entonces

$$T_G(x) = \mathbb{R}^n \quad \text{y} \quad N_G(x) = \{0\} \tag{3.46}$$

<u>*Prueba*</u> .- Sea $y \in \mathbb{R}^n$ arbitrario. Como $x \in \operatorname{int} G$, existe un $\delta > 0$ tal que $N(x;\delta) \subset G$. Escojamos sucesiones $x_j \to x$ y $t_j \to 0$ tales que

$$d_G^\circ(x;y) = \limsup_{j \to \infty} \frac{d_G(x_j + t_j y) - d_G(x_j)}{t_j}.$$

Entonces existe $j_0 \in \mathbb{N}$ tal que x_j, $x_j + t_j y \in B(x;\delta) \subset G$ para todo $j \geq j_0$. En consecuencia tendríamos que $d_G(x_j) = d_G(x_j + t_j y) = 0$ para todo $j \geq j_0$, entonces $d_G^\circ(x;y) = 0$ y en consecuencia, $y \in T_G(x)$. Ahora si $z \in N_G(x)$, entonces por definición tendríamos que

$$\langle y, z \rangle = 0 \quad \text{para todo} \quad y \in T_G(x) = \mathbb{R}^n$$

Aplicando esta propiedad para cualquier $y, -y \in \mathbb{R}^n$, tenemos que $z = 0$. ∎

Teorema 3.51. Si G_1 y $G_2 \subset \mathbb{R}^n$ son tales que $x \in G_1 \cap G_2$ y $x \in \operatorname{int} G_2$, entonces

$$T_{G_1}(x) = T_{G_1 \cap G_2}(x) \quad \text{y} \quad N_{G_1}(x) = N_{G_1 \cap G_2}(x) \tag{3.47}$$

<u>*Prueba*</u>.- Sea $y \in \mathbb{R}^n$ arbitrario. Supongamos que $x \in G_1 \cap G_2$ y $x \in \operatorname{int} G_2$. Entonces existe $\delta > 0$ tal que $B(x;\delta) \subset G_2$. Escojamos sucesiones $x_j^1, x_j^2 \to x$ y $t_j^1, t_j^2 \downarrow 0$ tales que

$$\begin{aligned} d^{\circ}_{G_1}(x;y) &= \limsup_{j\to\infty} \frac{d_{G_1}(x_j^1 + t_j^1 y) - d_{G_1}(x_j^1)}{t_j^1} \\ d^{\circ}_{G_1\cap G_2}(x;y) &= \limsup_{j\to\infty} \frac{d_{G_1\cap G_2}(x_j^2 + t_j^2 y) - d_{G_1\cap G_2}(x_j^2)}{t_j^2} \end{aligned}$$

Entonces existe $j_0 \in \mathbb{N}$ tal que $x_j^i,\ x_j^i + t_j^i y \in B(x;\delta) \subset G_2$ para $i = 1, 2$ y $j \geq j_0$. Ahora, el hecho de que $B(x;\delta) \cap G_1 = B(x;\delta) \cap G_1 \cap G_2$ implica

$$d_{G_1}|B(x;\delta) \cap G_1 = d_{G_1\cap G_2}|B(x;\delta) \cap G_1$$

Entonces, tomando $j \geq j_0$ tenemos

$$\begin{aligned} d^{\circ}_{G_1}(x;y) &= \limsup_{j\to\infty} \frac{d_{G_1}(x_j^1 + t_j^1 y) - d_{G_1}(x_j^1)}{t_j^1} \\ &= \limsup_{j\to\infty} \frac{d_{G_1\cap G_2}(x_j^1 + t_j^1 y) - d_{G_1\cap G_2}(x_j^1)}{t_j^1} \leq d^{\circ}_{G_1\cap G_2}(x;y) \end{aligned}$$

y

$$\begin{aligned} d^{\circ}_{G_1\cap G_2}(x;y) &= \limsup_{j\to\infty} \frac{d_{G_1\cap G_2}(x_j^2 + t_j^2 y) - d_{G_1\cap G_2}(x_j^2)}{t_j^2} \\ &= \limsup_{j\to\infty} \frac{d_{G_1}(x_j^2 + t_j^2 y) - d_{G_1}(x_j^2)}{t_j^2} \leq d^{\circ}_{G_1}(x;y) \end{aligned}$$

por consiguiente $d^{\circ}_{G_1}(x;y) = d^{\circ}_{G_1\cap G_2}(x;y)$ para todo $y \in \mathbb{R}^n$. Se concluye el teorema por las definiciones de cono tangente y cono normal. ∎

Capítulo 4

Soluciones, Métodos y Aplicaciones

4.1. Condiciones de Óptimo

Vamos a empezar dando las condiciones necesarias básicas para un problema de optimización sin restricciones y su mínimo x, que en el caso de funciones convexas, resultan ser también condiciones suficientes.

Teorema 4.1. Si $f : \mathbb{R}^n \mapsto \mathbb{R}$ es localmente Lipschitz en x y alcanza su mínimo en x, entonces

(i) $0 \in \partial f(x)$ y
(ii) $f^\circ(x; v) \geq 0 \quad \forall v \in \mathbb{R}^n$

<u>*Prueba*</u>.- Usando que f es localmente Lipschitz y que alcanza su minimo en x, tenemos:

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t} \geq \limsup_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t} \geq 0$$

y así

$$f^\circ(x; v) \geq 0 = \langle 0; v \rangle \quad \text{para todo} \quad v \in \mathbb{R}^n$$

lo que por definición siginifica que $0 \in \partial f(x)$. ∎

Teorema 4.2. Si $f : \mathbb{R}^n \to \mathbb{R}$ es convexa, entonces las siguientes condiciones son equivalentes:

(i) f alcanza su mínimo global en x,
(ii) $0 \in \partial_c f(x)$,
(iii) $f'(x; v) \geq 0 \quad \text{para todo} \quad v \in \mathbb{R}^n$

<u>*Prueba*</u>.- Supongamos (i), entonces $f(x) \leq f(y)$ para todo $y \in \mathbb{R}^n$, entonces

$$f(y) \geq f(x) = f(x) + \langle 0, y - x \rangle \quad \text{para todo} \quad y \in \mathbb{R}^n$$

lo que significa, por (3.2), que $0 \in \partial_c f(x)$.
Supongamos (ii) ahora. Entonces, sabiendo que $f'(x; v) = \text{máx}\{\langle \xi, v \rangle \mid \xi \in \partial_c f(x)\}$, tenemos que $f'(x; v) \geq \langle 0, v \rangle = 0$.
Finalmente, supongamos (iii). Sea $y \in \mathbb{R}^n$ arbitrario. Entonces,

$$f'(x; y - x) = \text{máx}\{\langle \xi, y - x \rangle \mid \xi \in \partial_c f(x)\}$$

Entonces, usando la hipótesis, existe un $\xi_y \in \partial_c f(x)$ tal que

$$0 \leq f'(x; y - x) = \langle \xi_y, y - x \rangle$$

Entonces por definición tendríamos que

$$f(y) \geq f(x) + \langle \xi_y, y - x \rangle \geq f(x)$$

que al haber tomado y arbitrario, significa que f alcanza su mínimo global en x. ∎

Recordando la teoría de ε-subdiferenciales, obtenemos la siguiente condición de óptimo.

Teorema 4.3. Si $f : \mathbb{R}^n \to \mathbb{R}$ es Lipschitz en una vecindad de x y alcanza su mínimo local en x, entonces

$$0 \in \partial_\varepsilon^G f(x) \tag{4.1}$$

<u>*Prueba*</u>.- Usando los teoremas 4.1 y 3.37, sale directamente. ∎

Teorema 4.4. Si $f : \mathbb{R}^n \to \mathbb{R}$ es convexa, entonces las siguientes condiciones son equivalentes:

(i) $0 \in \partial_\varepsilon f(x)$,
(ii) x minimiza f con ε

<u>*Prueba*</u>.- Supongamos que $0 \in \partial_\varepsilon f(x)$. Por definición de ε-subdiferencial, esto equivale a

$$f(y) \geq f(x) + \langle 0, y - x \rangle - \varepsilon = f(x) - \varepsilon \quad \text{para todo} \quad y \in \mathbb{R}^n$$

que, en otras palabras, significa que x minimiza a f con ε. ∎

Para problemas restringidos necesitamos el siguiente lema.

Lema 4.5. Sea f Lipschitz con constante K en el conjunto S. Sea $x \in G \subset S$ y supongamos que f alcanza su mínimo sobre G en x. Entonces para cualquier $\widehat{K} \geq K$, la función $g(y) = f(y) + \widehat{K}\mathrm{d}_G(y)$ alcanza su mínimo sobre S en x. Si $\widehat{K} > K$ y G es cerrado, entonces cualquier minimizador de g sobre S debe estar en G.

Prueba.- Probemos lo primero por contradicción. Entonces debe haber un punto $y \in S$ y $\varepsilon > 0$ tal que $g(y) < g(x) - \widehat{K}\varepsilon$, esto es

$$g(y) = f(y) + \widehat{K}\mathrm{d}_G(y) < f(x) + \widehat{K}\mathrm{d}_G(x) - \widehat{K}\varepsilon = f(x) - \widehat{K}\varepsilon$$

Sea $c \in G$ tal que $\|y - c\| \leq \mathrm{d}_G(y) + \varepsilon$. Entonces por la condición de Lipschitz tenemos

$$\begin{aligned} f(c) &\leq f(y) + K\|y-c\| \leq f(y) + \widehat{K}\|y-c\| \\ &\leq f(y) + \widehat{K}(\mathrm{d}_G(y) + \varepsilon) \\ &< f(x) + \widehat{K}\varepsilon - \widehat{K}\varepsilon = f(x) \end{aligned}$$

lo que contradice que x minimize f en G.
Supongamos ahora que $\widehat{K} > K$ y que G es cerrado; y que y también minimiza g sobre S. Entonces, podemos aplicar la primera afirmación a la constante $\frac{\widehat{K}+K}{2}$ y obtener

$$f(y) + \widehat{K}\mathrm{d}_G(y) = f(x) = f(x) + (\widehat{K} + K)\mathrm{d}_G(x)/2 \leq f(y) + (\widehat{K} + K)\mathrm{d}_G(y)/2$$

lo que implica que $\widehat{K}\mathrm{d}_G(y) \leq (\widehat{K} + K)\mathrm{d}_G(y)/2$, que sólo podría ser cierto si es que $\mathrm{d}_G(y) = 0$. Como G es cerrado, esto implica que $y \in G$. ■

Ahora veamos las condiciones cuando los problemas son restringidos.

Teorema 4.6. Si f es Lipschitz en una vecindad de x y alcanza su mínimo local sobre el conjunto $G \subset \mathbb{R}^n$ en x, entonces

$$0 \in \partial f(x) + N_G(x) \tag{4.2}$$

Prueba.- Sea $S \subset \mathbb{R}^n$ un conjunto abierto que contiene a x, y supongamos que f es Lipschitz con constante K en S. Entonces, $G \cap S \subset S$ y f alcanza su mínimo en $G \cap S$ en x. Entonces, por el lema 4.5, la función $f(y) + K\mathrm{d}_{G\cap S}(y)$ alcanza su mínimo sobre S en x y por el teorema 4.1(i), tenemos que

$$0 \in \partial(f + K\mathrm{d}_{G\cap S})(x)$$

De los teoremas 3.49 y 3.51, obtenemos el resultado buscado

$$\begin{aligned} 0 \in \partial(f + K\mathrm{d}_{G\cap S})(x) &\subset \partial f(x) + K\partial \mathrm{d}_{G\cap S}(x) \\ &\subset \partial f(x) + N_{G\cap S}(x) = \partial f(x) + N_G(x) \end{aligned}$$

■

Teorema 4.7. Si $f : \mathbb{R}^n \to \mathbb{R}$ es convexa, entonces las siguientes condiciones son equivalentes:

(i) $0 \in \partial_c f(x) + N_G(x)$

(ii) f alcanza su mínimo global sobre G en x.

<u>*Prueba*</u>.- Del Teorema 4.6, vemos que (ii) implica (i). Supongamos ahora que $0 \in \partial_c f(x)+N_G(x)$. Entonces existe un $\xi \in \partial_c f(x)$ y $z \in N_G(x)$ tal que $\xi = -z$. Por la definición de subdiferencial para funciones convexas, tenemos que

$$f(y) \geq f(x) + \langle \xi, y-x\rangle = f(x) - \langle z, y-x\rangle \quad \text{para todo} \quad y \in N_G(x)$$

y, sabiendo que el cono normal tambien puede ser escrito como

$$N_G(x) = \{z \in \mathbb{R}^n \mid \langle y-x, z\rangle \leq 0 \quad \text{para todo} \quad y \in G\}$$

se tiene que $\langle z, y-x\rangle \leq 0$, y asi

$$f(y) \geq f(x) - \langle z, y-x\rangle \geq f(x) \quad \text{para todo} \quad y \in G$$

■

Es frecuente ver en la práctica que el conjunto de restricciones G tiene una forma especial.

Corolario 4.8. Sea $F : \mathbb{R}^n \to \mathbb{R}$ una función tal que $F(y) = \max\{F_i(y) \mid i = 1, \ldots, m\}$ donde cada $F_i : \mathbb{R}^n \to \mathbb{R}$ es Lipschitz localmente en x y se cumple o que $F(x) < 0$ o que $0 \notin \partial F(x)$; y

$$G = \{y \in \mathbb{R}^n \mid F(y) \leq 0\}. \tag{4.3}$$

Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es Lipschitz localmente en x y alcanza su mínimo local sobre G en x. Entonces, existe $\lambda_i \geq 0$ para $i = 1, \ldots, m$ tal que $\lambda_i F_i(x) = 0$ y

$$0 \in \partial f(x) + \sum_{i=1}^{m} \lambda_i \partial F_i(x) \tag{4.4}$$

<u>*Prueba*</u>.- Evidentemente, la función F es Lipschitz localmente en x. Si $F(x) < 0$, entonces $x \in \mathrm{int}G$, y asi, por el teorema 3.50, $N_G(x) = \{0\}$. Entonces por el teorema 4.6 se tiene q que $0 \in \partial f(x)$, entonces podemos tomar $\lambda_i = 0$ para cada $i = 1, \ldots, m$.

Supongamos ahora que $F(x) = 0$, entonces el conjunto $I(x) = \{i \in \{1,\ldots,m\} \mid F_i(x) = F(x)\}$ es no vacío. Entonces podemos escribir el conjunto G de la forma

$$G = \{y \in \mathbb{R}^n \mid (y, F(x)) \in \text{epi } F\}.$$

Como $0 \notin \partial F(x)$, implica que (ver [Makela-1990])

$$N_G(x) \subset \bigcup_{\lambda \geq 0} \lambda \partial F(x)$$

Por otro lado tenemos que

$$\partial F(x) \subset \text{conv}\,\{\partial F_i(x) \mid i \in I(x)\},$$

entonces por el teorema 4.6, existen $\lambda_i \geq 0$ tales que $\lambda_i F_i(x) = 0$ y

$$0 \in \partial f(x) + \sum_{i=1}^{m} \lambda_i \partial F_i(x) \qquad \blacksquare$$

Corolario 4.9. Sea $F : \mathbb{R}^n \to \mathbb{R}$ una función tal que $F(y) = \text{máx}\{F_i(y) \mid i = 1,\ldots,m\}$ donde $F_i : \mathbb{R}^n \to \mathbb{R}$ es convexa y $F(z) < 0$ para algún $z \in \mathbb{R}^n$. Sea

$$G = \{y \in \mathbb{R}^n \mid F(y) \leq 0\}$$

Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es convexa. Entonces las siguientes condiciones son equivalentes:

(i) f alcanza su mínimo global sobre G en x,

(ii) existen $\lambda_i \geq 0$ para $i = 1,\ldots,m$ tales que $\lambda_i F_i(x) = 0$ y

$$0 \in \partial_c f(x) + \sum_{i=1}^{m} \lambda_i \partial_c F_i(x) \tag{4.5}$$

Prueba.- Por la definición de F, se ve fácilmente que es convexa. Como existe un $z \in \mathbb{R}^n$ tal que $F(z) < 0$, vemos que si $F(x) = 0$, entonces x no podría ser mínimo global de F sobre G. Entonces, por el teorema 4.2 tenemos que $0 \notin \partial_c F(x)$ y el corolario 4.8 implica que (ii) es implicado por (i).

Supongamos ahora que se cumple (ii). Entonces

$$\sum_{i=1}^{m} \lambda_i \partial_c F_i(x) \subset \text{conv}\,\{\partial_c F_i(x) \mid i \in I(x)\}$$

Como F es una combinación lineal de funciones convexas, F es regular, entonces

$$\text{conv}\,\{\partial_c F_i(x) \mid i \in I(x)\} = \partial_c F(x)$$

lo que implica que

$$\partial_c F(x) \subset \bigcup_{\lambda \geq 0} \lambda \partial_c F(x) = N_G(x)$$

Entonces tenemos que

$$0 \in \partial_c f(x) + N_G(x)$$

lo que por el teorema 4.7 implica (i). ∎

4.2. Linearización de Problemas no Restringidos

En esta sección vamos a definir algunas nociones de linearización para funciones Lipschitz y presentar algunas de sus propiedades básicas. Con estas linearizaciones vamos a poder construir después aproximaciones lineales por partes de problemas de optimización no restringidos.

Definición 4.10. Supongamos que la función $f : \mathbb{R}^n \to \mathbb{R}$ es Lipschitz cerca de $x \in \mathbb{R}^n$ y sea $\xi \in \partial f(x)$ arbitrario. Entonces la ξ-linearización de f en x es la función $\hat{f}_\xi : \mathbb{R}^n \to \mathbb{R}$ definida por

$$\hat{f}_\xi(y) = f(x) + \langle \xi, y - x \rangle \quad \text{para todo} \quad y \in \mathbb{R}^n \tag{4.6}$$

y la *linearización de f en x* es la función $\hat{f}_x : \mathbb{R}^n \to \mathbb{R}$ tal que

$$\hat{f}_x(y) = \text{máx}\{\hat{f}_\xi(y) \mid \xi \in \partial f(x)\} \quad \text{para todo} \quad y \in \mathbb{R}^n \tag{4.7}$$

Vamos a mostrar algunos hechos elementales de estas linearizaciones.

Teorema 4.11. Sea $f : \mathbb{R}^n \to \mathbb{R}$ de Lipschitz cerca a x.
Entonces la linearización $\hat{f}_x$ es convexa y

(i) $\hat{f}_x(x) = f(x)$,
(ii) $\hat{f}_x(y) = f(x) + f^{\circ}(x; y - x) \quad$ para todo $\quad y \in \mathbb{R}^n$,
(iii) $\partial \hat{f}_x(x) = \partial f(x)$

Prueba.- Vamos a mostrar primero que $\hat{f}_x$ es convexa. Sean $y, z \in \mathbb{R}^n$ y $\lambda \in [0, 1]$. Entonces

$$\begin{aligned}
\hat{f}_x(\lambda y + (1-\lambda) z) & \quad \text{máx}\{f(x) + \langle \xi, \lambda y + (1-\lambda) z - x \rangle \mid \xi \in \partial f(x)\} \\
& \leq \lambda \,\text{máx}\{f(x) + \langle \xi, y - x \rangle \mid \xi \in \partial f(x)\} \\
& + (1-\lambda) \,\text{máx}\{f(x) + \langle \xi, z - x \rangle \mid \xi \in \partial f(x)\} \\
& = \lambda \hat{f}_x(y) + (1-\lambda) \hat{f}_x(z)
\end{aligned}$$

Por definición tenemos que

$$\hat{f}_x(x) = \text{máx}\{f(x) + \langle \xi, x - x\rangle \mid \xi \in \partial f(x)\} = f(x)$$

lo que demuestra (i). Para demostrar (ii) hacemos lo siguiente:

$$\begin{aligned}
\hat{f}_x(y) &= \text{máx}\{f(x) + \langle \xi, y - x\rangle \mid \xi \in \partial f(x)\} \\
&= f(x) + \max_{\xi \in \partial f(x)} \langle \xi, y - x\rangle \\
&= f(x) + f^\circ(x; y - x) \quad \text{para todo} \quad y \in \mathbb{R}^n
\end{aligned}$$

Para (iii), sea $\xi \in \partial f(x)$. Entonces por (i) tenemos que

$$\hat{f}_x(y) \geq \hat{f}_\xi(y) = f(x) + \langle \xi, y - x\rangle = \hat{f}_x(x) + \langle \xi, y - x\rangle$$

entonces $\xi \in \partial_c \hat{f}_x(x) = \partial \hat{f}_x(x)$. Por otro lado, por (ii) tenemos

$$\begin{aligned}
\hat{f}_x^\circ(x; v) &= \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} \frac{\hat{f}_x(y + \lambda v) - \hat{f}_x(y)}{\lambda} \\
&= \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} \frac{f(x) + f^\circ(x; y + \lambda v - x) - f(x) - f^\circ(x; y - x)}{\lambda}
\end{aligned}$$

como f° es positivamente homogénea y subaditiva

$$\begin{aligned}
\hat{f}_x^\circ(x; v) &\leq \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} \frac{f^\circ(x; y - x) + f^\circ(x, \lambda v) - f^\circ(x; y - x)}{\lambda} \\
&= \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} \frac{\lambda f^\circ(x; v)}{\lambda} \\
&= f^\circ(x; v),
\end{aligned}$$

entonces $\partial \hat{f}_x(x) \subset \partial f(x)$, lo que completa la prueba. ∎

Si la función f resulta ser convexa, entonces obtenemos el siguiente útil resultado.

Teorema 4.12. Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Entonces,

1. $f(y) = \text{máx}\{\hat{f}_x(y) \mid x \in \mathbb{R}^n\}$ para todo $y \in \mathbb{R}^n$,
2. $\hat{f}_x(y) \leq f(y)$ para todo $y \in \mathbb{R}^n$,
3. epi $f \subset$ epi $\hat{f}_x$

<u>*Prueba*</u>.- La convexidad de f implica que para todo $y \in \mathbb{R}^n$ tenemos

$$\begin{aligned} f(y) &= \max\{f(x) + \langle \xi, y - x\rangle \mid \xi \in \partial f(x),\ x \in \mathbb{R}^n\} \\ &= \max\{\hat{f}_\xi(y) \mid \xi \in \partial f(x),\ x \in \mathbb{R}^n\} \\ &= \max\{\hat{f}_x(y) \mid x \in \mathbb{R}^n\}. \end{aligned}$$

Las afirmaciones (ii) y (iii) se cumplen directamente por (i). ■

Un problema fundamental en los métodos de optimización iterativa es encontrar una dirección para la cual los valores de la función decrece si se mueve por esa dirección. Por esta razón vamos a necesitar la siguiente definición.

<u>Definición</u> 4.13. La dirección $d \in \mathbb{R}^n$ se dice que es una *dirección de descenso* para $f : \mathbb{R}^n \to \mathbb{R}$ en x si existe un $\varepsilon > 0$ tal que

$$f(x + td) < f(x) \quad \text{para todo} \quad t \in \langle 0, \varepsilon] \tag{4.8}$$

Vamos a ver cómo obtener direcciones de descenso para funciones Lipschitz.

Teorema 4.14. Sea $f : \mathbb{R}^n \to \mathbb{R}$ de Lipschitz cerca a x. La dirección $d \in \mathbb{R}^n$ es una dirección de descenso para f en x si se cumple cualquiera de las siguientes proposiciones:

1. $f^\circ(x; d) < 0$,
2. $\langle \xi, d\rangle < 0 \quad \text{para todo} \quad \xi \in \partial f(x)$,
3. $\langle \xi, d\rangle < 0 \quad \text{para todo} \quad \xi \in \partial^G_\varepsilon f(x)$,
4. d es una dirección de descenso para $\hat{f}_x$ en x

<u>*Prueba*</u>.- Por la definición de derivada direccional generalizada,

$$\limsup_{t\downarrow 0} \frac{f(x + td) - f(x)}{t} \leq \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + td) - f(y)}{t} = f^\circ(x; d) < 0$$

Entonces, por la definición de límite superior, existe $\varepsilon > 0$ tal que $f(x + td) - f(x) < 0$ para todo $t \in \langle 0, \varepsilon]$, lo que significa que d es una dirección de descenso para f en x.
La parte (ii) se obtiene de (i) y usando el resultado que

$$f^\circ(x; d) = \max\{\langle \xi, d\rangle \mid \xi \in \partial f(x)\} < 0$$

La parte (iii) sigue de la parte (ii), ya que

$$\partial f(x) \subset \partial^G_\varepsilon f(x)$$

Para probar la afirmación (iv), sea d un dirección de descenso para $\hat{f}_x$ en x. Entonces existe $\varepsilon > 0$ tal que

$$\hat{f}_x(x + td) < \hat{f}_x(x) \quad \text{para todo} \quad t \in \langle 0, \varepsilon]$$

Entonces, por el teorema 4.11 y usando el hecho que $f^\circ(x; v)$ es positivamente homogénea,

$$0 > f(x) + f^\circ(x; x + td - x) - f(x) = t \cdot f^\circ(x; d) \quad \text{para todo} \quad t \in \langle 0, \varepsilon]$$

y usando (i) se completa la prueba. ■

Ahora con los siguientes dos lemas vamos a demostrar el resultado principal de este trabajo.

Lema 4.15. Sea $G \subset \mathbb{R}^n$ un conjunto compacto convexo. Entonces para $p \in G$

$$p = \arg\min\{\|g\| \mid g \in G\} \quad \Longleftrightarrow \quad \langle p, g\rangle \geq \|p\|^2 \quad \text{para todo} \quad g \in G \tag{4.9}$$

<u>*Prueba*</u>.- Vamos a demostrar primero que $\arg\min\{\|g\| \mid g \in G\}$ está bien definido. Como la función $\|\cdot\|$ es continua, alcanza su mínimo sobre el conjunto compacto G. Si $0 \in G$, ese es el único minimizador. Supongamos que existen dos puntos $p_1, p_2 \in G$ distintos y no nulos, minimizadores de g. Sea $\lambda \in \langle 0, 1\rangle$, entonces por la convexidad de G, el punto $\lambda p_1 + (1 - \lambda)p_2$ está en G. Si p_1 y p_2 son linealmente independientes, tenemos

$$\|\lambda p_1 + (1 - \lambda)p_2\| < \lambda\|p_1\| + (1 - \lambda)\|p_2\| = \arg\min\{\|g\| \mid g \in G\}$$

lo que es una contradicción. Por otro lado, si p_1 y p_2 son linealmente dependientes, entonces existe $\alpha \neq 1$ tal que $p_2 = \alpha p_1$. Si $\alpha \leq 0$, entonces por la convexidad de G, $0 \in G$ lo que no puede ser, y si $\alpha > 0$, se tiene

$$\|p_2\| = \alpha\|p_1\| \neq \|p_1\|,$$

lo que es una contradicción dado que los dos puntos son puntos mínimos. Por esto, el mínimo es único y $\arg\min\{\|g\| \mid g \in G\}$ está bien definido.

Continuamos ahora, supongamos primero que $\langle p, g\rangle \geq \|p\|^2$ para todo $g \in G$. Entonces tenemos,

$$\|p\|^2 \leq \langle p, g\rangle \leq \|p\|\|g\| \quad \text{para todo} \quad g \in G$$

por lo cual

$$\|p\| \leq \arg\min\{\|g\| \mid g \in G\}$$

entonces, $p = \arg\min\{\|g\| \mid g \in G\}$. Supongamos ahora que $p = \arg\min\{\|g\| \mid g \in G\}$ y que existe $g \in G$ tal que $\langle g, p\rangle < \|p\|^2$. Como G es convexo, el punto $\lambda g + (1-\lambda)p$ está en G para cada $\lambda \in \langle 0, 1\rangle$ y

$$\begin{aligned}
\|\lambda g + (1-\lambda)p\|^2 &= \|p + \lambda(g-p)\|^2 \\
&= \langle p + \lambda(g-p), p + \lambda(g-p)\rangle \\
&= \|p\|^2 + 2\lambda[\langle g, p\rangle - \|p\|^2] + \lambda^2\|g-p\|^2 \\
&< \|p\|^2,
\end{aligned}$$

cuando λ es lo suficientemente pequeño. Esto contradice que p sea el punto mínimo, entonces $\langle g, p\rangle \geq \|p\|^2$ para todo $g \in G$. ∎

Lema 4.16. Si $f : \mathbb{R}^n \to \mathbb{R}$ es Lipschitz cerca de x y $d \in \mathbb{R}^n$ es una dirección arbitraria, entonces la función $f^\circ(x;\cdot)$ es Lipschitz cerca de d y

$$\partial f^\circ(x;d) \subset \{\xi \in \partial f(x) \mid \langle \xi, d\rangle = f^\circ(x;d)\}. \tag{4.10}$$

<u>*Prueba*</u>.- Por el teorema 3.32(ii), la función $f^\circ(x;\cdot)$ es Lipschitz cerca de d. Supongamos ahora que $\xi \in \partial f^\circ(x;d)$, entonces por el teorema 3.32(i) tenemos que para todo $v \in \mathbb{R}^n$

$$\begin{aligned}
\langle \xi, v\rangle &\leq (f^\circ)^\circ((x;d);v) - \limsup_{\substack{d' \to d \\ \lambda \downarrow 0}} \frac{f^\circ(x;d'+\lambda v) - f^\circ(x;d')}{\lambda} \\
&\leq \limsup_{\substack{d' \to d \\ \lambda \downarrow 0}} \frac{f^\circ(x;d') + \lambda f^\circ(x;v) - f^\circ(x;d')}{\lambda} = \limsup_{\substack{d' \to d \\ \lambda \downarrow 0}} f^\circ(x;v) \\
&= f^\circ(x;v),
\end{aligned}$$

entonces $\xi \in \partial f(x)$. De lo anterior, si reemplazamos d en lugar de v, tenemos

$$\begin{aligned}
f^\circ(x;d) &\geq (f^\circ)^\circ((x;d);d) \\
&= \limsup_{\substack{d' \to d \\ \lambda \downarrow 0}} \frac{f^\circ(x;d'+\lambda d) - f^\circ(x;d')}{\lambda} \\
&\geq \limsup_{\substack{d' \to d \\ \lambda \downarrow 0}} \frac{f^\circ(x;(1+\lambda)d) - f^\circ(x;d)}{\lambda} \\
&= \limsup_{\substack{d' \to d \\ \lambda \downarrow 0}} \frac{f^\circ(x;d) + \lambda f^\circ(x;d) - f^\circ(x;d)}{\lambda} = \limsup_{\substack{d' \to d \\ \lambda \downarrow 0}} f^\circ(x;d) \\
&= f^\circ(x;d),
\end{aligned}$$

entonces, $f^{\circ}(x;d) = (f^{\circ})^{\circ}((x;d);d)$. Si tenemos que $\langle \xi, d\rangle < f^{\circ}(x;d)$, entonces

$$-(f^{\circ})^{\circ}((x;d);-d) \leq -\langle \xi, -d\rangle = \langle \xi, d\rangle < f^{\circ}(x;d) = (f^{\circ})^{\circ}((x;d);d)$$

y entonces

$$\begin{aligned}
&\liminf_{\substack{d'\to d\\ \lambda\downarrow 0}} \frac{f^{\circ}(x;d'+\lambda d) - f^{\circ}(x;d')}{\lambda} = -\limsup_{\substack{d'\to d\\ \lambda\downarrow 0}} \frac{f^{\circ}(x;d') - f^{\circ}(x;d'+\lambda d)}{\lambda} = \\
&-\limsup_{\substack{d'\to d\\ \lambda\downarrow 0}} \frac{f^{\circ}(x;d'+\lambda d-\lambda d) - f^{\circ}(x;d'+\lambda d)}{\lambda} = -(f^{\circ})^{\circ}((x;d);-d) \\
&< (f^{\circ})^{\circ}((x;d);d) = \limsup_{\substack{d'\to d\\ \lambda\downarrow 0}} \frac{f^{\circ}(x;d'+\lambda d) - f^{\circ}(x;d')}{\lambda}
\end{aligned}$$

entonces

$$\liminf_{\substack{d'\to d\\ \lambda\downarrow 0}} \frac{f^{\circ}(x;d'+\lambda d) - f^{\circ}(x;d')}{\lambda} < \limsup_{\substack{d'\to d\\ \lambda\downarrow 0}} \frac{f^{\circ}(x;d'+\lambda d) - f^{\circ}(x;d')}{\lambda}$$

Esto significa que no existe la derivada direccional $(f^{\circ})'(x;d)$. Por otro lado, por el teorema 3.32(i), la función $v \mapsto f^{\circ}(x;v)$ es positivamente homogénea y subaditiva, entonces es convexa, y por el teorema 3.13(ii), esta función es regular, lo que contradice que no exista la derivada direccional $(f^{\circ})'(x;d)$. Entonces tenemos que

$$\langle \xi, d\rangle = f^{\circ}(x;d)$$

completando así la prueba. ∎

La mayoría de los métodos de optimización no suave están basados en el siguiente teorema, que dice cómo es que podemos encontrar una dirección de descenso para la función de linearización. Está dirección, según el teorema 4.14(iv) también sería una dirección de descenso para la función original.

Teorema 4.17. Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es Lipschitz cerca de x y $\xi^* \in \partial f(x)$ tal que $\xi^* = \arg\min\{\|\xi\| \mid \xi \in \partial f(x)\}$. Considere el problema

$$\min \hat{f}_x(x+d) + \tfrac{1}{2}\|d\|^2 \quad \text{para todo} \quad d \in \mathbb{R}^n, \tag{4.11}$$

Entonces

(i) El problema (4.11) tiene una solución única $d^* \in \mathbb{R}^n$ tal que $-d^* = \xi^* \in \partial f(x)$,

(ii) $f^{\circ}(x;d^*) = -\|d^*\|^2$,

(iii) $\hat{f}_x(x + \lambda d^*) = \hat{f}_x(x) - \lambda\|\xi^*\|^2$ para todo $\lambda \in [0, 1]$,
(iv) $0 \notin \partial f(x) \iff d^* \neq 0$,
(v) $0 \in \partial f(x) \iff \hat{f}_x$ alcanza su mínimo global en x.

Prueba.- Vamos a empezar demostrando las afirmaciones (i) y (ii). Definamos una función $\rho : \mathbf{R}^n \to \mathbf{R}$ tal que para todo $d \in \mathbf{R}^n$

$$\rho(d) = \hat{f}_x(x + d) + \tfrac{1}{2}\|d\|^2 = f(x) + f^\circ(x; d) + \tfrac{1}{2}\|d\|^2$$

Veamos que esta función ρ es estrictamente convexa. Sean $d, d' \in \mathbf{R}^n$ y $\lambda \in [0, 1]$. Entonces por el teorema 3.32(i) y el hecho de que el mapeo $t \mapsto t^2$ es estrictamente convexo, tenemos

$$\begin{aligned}
\rho(\lambda d + (1-\lambda)d') &= f(x) + f^\circ(x; \lambda d + (1-\lambda)d') + \tfrac{1}{2}\|\lambda d + (1-\lambda)d'\|^2 \\
&< f(x) + \lambda f^\circ(x; d) + (1-\lambda)f^\circ(x; d') + \tfrac{1}{2}\lambda^2\|d\|^2 + \tfrac{1}{2}(1-\lambda)^2\|d'\|^2 \\
&< \lambda(f(x) + f^\circ(x; d) + \tfrac{1}{2}\|d\|^2) + (1-\lambda)(f(x) + f^\circ(x; d') + \tfrac{1}{2}\|d'\|^2) \\
&= \lambda\rho(d) + (1-\lambda)\rho(d')
\end{aligned}$$

Como la función ρ es estrictamente convexa y para cualquier $\xi \in \partial f(x)$ tenemos

$$\rho(d) \geq f(x) + \langle \xi, d \rangle + \tfrac{1}{2}\|d\|^2 \geq f(x) - \|\xi\|\|d\| + \tfrac{1}{2}\|d\|^2 \longrightarrow \infty$$

cuando $\|d\| \to \infty$, existe un único $d^* \in \mathbf{R}^n$ que minimiza a ρ. Entonces por el teorema 4.2 tenemos que $0 \in \partial\rho(d^*)$. El hecho de que las funciones $d \mapsto f^\circ(x; d)$ y $d \mapsto \frac{1}{2}\|d\|^2$ son convexas y por lo mismo regulares implica que

$$\partial\rho(d^*) = \partial(f^\circ(x; d^*) + \tfrac{1}{2}\|d^*\|^2) = \partial f^\circ(x; d^*) + \partial(\tfrac{1}{2}\|d^*\|^2),$$

La función $d \mapsto \frac{1}{2}\|d\|^2$ es también continuamente diferenciable, entonces el único elemento de su subdiferencial es la gradiente [Makela-1990], por lo que

$$\partial(\tfrac{1}{2}\|d^*\|^2) = d^*$$

y el lema 4.16 implica que

$$\partial f^\circ(x; d^*) \subset \{\xi \in \partial f(x) \mid \langle \xi, d^* \rangle = f^\circ(x; d^*)\}.$$

Entonces tenemos

$$0 \in d^* + \{\xi \in \partial f(x) \mid \langle \xi, d^* \rangle = f^\circ(x; d^*)\}$$

lo que significa que existe un $\hat{\xi} \in \partial f(x)$ tal que $\hat{\xi} = -d^*$ y $f^{\circ}(x; d^*) = -\|\hat{\xi}\|^2$.
El hecho de que $\hat{\xi} \in \partial f(x)$, y que

$$\|\hat{\xi}\|^2 = -f^{\circ}(x; d^*) \leq -\langle \xi, d^* \rangle = \langle \xi, \hat{\xi} \rangle$$

para todo $\xi \in \partial f(x)$ implican, por el lema 4.15, que $\hat{\xi} = \arg \min\{\|\xi\| \mid \xi \in \partial f(x)\} = \xi^*$, lo que establece las afirmaciones (i) y (ii).
La afirmación (iii) sigue de los teoremas 4.11, 3.1.2(i) (Makela) y de la afirmación (ii) de este teorema ya que

$$\hat{f}_x(x + \lambda d^*) = f(x) + f^{\circ}(x; \lambda d^*) = \hat{f}_x(x) + \lambda f^{\circ}(x; d^*) = \hat{f}_x(x) - \lambda \|\xi^*\|^2$$

Para demostrar (iv), supongamos que $0 \notin \partial f(x)$. Esto es equivalente a

$$\|d^*\| = \|\xi^*\| = \min_{\xi \in \partial f(x)} \|\xi\| > 0$$

lo que equivale a decir que $d^* \neq 0$. La afirmación (v) sigue directamente de los teoremas 4.11 y 4.2(ii). ∎

4.3. Aplicación

Una aplicación que se le da a esta teoría es solucionar problemas de *Equilibrio de Transporte*. Vamos a plantear el problema y ver como es modelado como una Desigualdad Variacional.

Equilibrio de Transporte

El problema se da en una Red de Transporte, que es representada por un *Grafo Dirigido* $\mathbb{G}(\mathcal{N}, \mathcal{L})$, donde los conjuntos que vamos a usar son

$\mathcal{N}$: Conjunto de Nodos
$\mathcal{L}$: Conjunto de Arcos Dirigidos
$\mathcal{I}$: Conjunto de todos los pares origen-destino conectados

Los conjuntos $\mathcal{N}, \mathcal{L}, \mathcal{I}$ tienen N, L, I elementos respectivamente

A cada arco $l \in \mathcal{L}$ se le asocia una función de costo de viaje sobre el arco: $C_l(f)$, que es una función continua bien definida que depende de la función flujo sobre los arcos, $f \in \mathbb{R}^L$.

Observación 4.18. En muchos de los modelos, esta función sólo depende del flujo sobre el mismo arco, lo cual excluye la posibilidad de representar interacciones entre diferentes flujos sobre arcos.

En una red de transporte, todos los viajes se originan y terminan en centroides (nodos). Asi

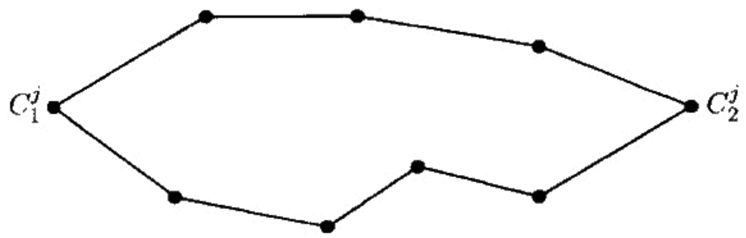
$$\begin{array}{rcl} \mathcal{J} & : & \text{Conjunto de todos los pares de centroides} \\ & & \text{El conjunto } \mathcal{J} \text{ tiene } J \text{ elementos} \end{array}$$

Para cada par de centroides $j \in \mathcal{J}$, existe un subconjunto $\mathcal{I}_j \in \mathcal{I}$ de viajes (a través de las rutas) entre ese par de centroides j.

Ejemplo 4.19. $j = (C_1^j, C_2^j)$

Modelo:

$$\min_f \sum_{l \in \mathcal{L}} C_l(f) \cdot f_l$$
$$s.a.(restricciones)$$



Modelo de Elección de Rutas

Consiste en definir un criterio, una hipótesis, con el cual se pueda elegir la ruta a seguir en la red de transporte y hacer una caracterización matemática de ella.

Notación:

$\mathcal{P}_i$: Conjunto de rutas entre el par i O/D, $i \in \mathcal{I}$

$\mathcal{P} = \bigcup_{i \in \mathcal{I}} \mathcal{P}_i$: $\mathcal{P}_i$ y $\mathcal{P}$ contienen P_i y P rutas totales respectivamente

$h \in \mathbb{R}^P$ denota el vector de flujo sobre las rutas, y se dice que es factible para un vector de elección de viaje (demanda para viajar en el par O/D) $d \in \mathbb{R}^I$ dado, si

$$\sum_{r \in \mathcal{P}_i} h_r = d_i \qquad \forall i \in \mathcal{I} \text{ (Satisfacción de la demanda)} \tag{4.12}$$

$$h_r \geq 0 \qquad \forall r \in \mathcal{P} \tag{4.13}$$

Dada esta notación, vamos a llamar a $\mathbf{H}(d) \subset \mathbb{R}^P$ al conjunto de flujos de caminos factibles. Para cualquier $h \in \mathbf{H}(d)$, existe un correspondiente flujo de arco factible $f \in \mathbb{R}^L$ obtenido de

$$f = Ah \tag{4.14}$$

donde $A = [a_{qr}]$ es una matriz de incidencia arco/camino para la red y

$$a_{qr} = \begin{cases} 1 & : \text{ si el camino } r \text{ usa el arco } q \\ 0 & : \text{ en otro caso} \end{cases}$$

Con esta defininción, $\mathbf{F}(d) \in \mathbb{R}^L$ es el conjunto de flujos de arco factibles dado un vector $d \in \mathbb{R}^I$.

Observación 4.20. Dada la definición, $\mathbf{H}(d)$ y $\mathbf{F}(d)$ son conjuntos convexos y compactos.

Para cualquier flujo factible h, su vector de costo de ruta $\mathbf{C}(h) \in \mathbb{R}^P$ puede ser obtenido a partir del vector de costo de arcos $C(f)$ como

$$\mathbf{C}(h) = A^T C(f) \tag{4.15}$$

donde f y h están relacionados por (4.14)

Observación 4.21. Inicialmente se asume que los usuarios escogen las rutas de acuerdo al principio óptimo de usuario de Wardrop, en el que ningún usuario reduce su costo conectándose unilateralmente a otra ruta. En otras palabras, las rutas usadas entre un par dado O/D tienen iguales costos de viaje y las rutas no usadas tienen un costo mayor.

Vamos a definir el flujo de camino que optimiza el uso de la red de transporte en cuanto al costo.

Definición 4.22. El flujo de camino h^* se dice que es una solución de equilibrio óptimo-usuario en la red, si satisface la siguiente condiciones (las cuales formalizan el principio de Wardrop en la elección de ruta por usuario): Hallar h^* que satisface lo siguiente:

$$\begin{aligned} C_r(h^*) - \mu_i = 0 \quad &, \quad \text{si } h_r^* > 0 \quad \text{para} \quad r \in \mathcal{P}_i, i \in \mathcal{I} \\ C_r(h^*) - \mu_i \geq 0 \quad &, \quad \text{si } h_r^* = 0 \quad \text{para} \quad r \in \mathcal{P}_i, i \in \mathcal{I} \\ h^* &\in \mathbf{H}(d) \\ \mu &\geq 0 \end{aligned} \tag{4.16}$$

donde $\mu \in \mathbb{R}^I$ y μ_i es el costo de viaje <u>igual</u> para todos las <u>rutas usadas</u> entre cada par i O/D (en el equilibrio)

<u>Modelo de Elección de Viaje</u>

Asumimos que la forma funcional del modelo que representa las elecciones de los usuarios de pares O/D para viajar ha sido especificada. Designamos por:

$$\begin{aligned} \mathbf{H} : \mathbb{R}^I &\longrightarrow \mathbb{R}^I \\ \mu &\longrightarrow \mathbf{H}(\mu) = \begin{pmatrix} H_1(\mu) \\ \vdots \\ H_I(\mu) \end{pmatrix} \end{aligned} \qquad \text{La elección de viaje está dada por} \quad d_i = H_i(\mu) \quad \forall i \in \mathcal{I} \tag{4.17}$$

H también contiene otras variables y parámetros explicativos, los cuales, ya que son independientes del flujo, no son representados explícitamente en (4.17).

<u>Observación</u> 4.23. Algunos modelos de elección de viaje deben satisfacer condiciones adecuadas expresadas como restricciones. Un ejemplo típico es aquel problema que usa el origen y/o destino como restricción en modelos de intersección espacial.

4.3.1. Formulación como Desigualdad Variacional de Modelos Combinados

Este modelo es la combinación de los modelos (4.16) y (4.17) en un solo, en el cual debe cumplirse las condiciones de elección de ruta (4.16) y las condiciones de elección de viaje (4.17) simultáneamente; y se consigue expresando todo en una Desigualdad Variacional.

Aashtiani y Magnanti han demostrado que este problema se puede expresar como un problema del tipo NCP (Problema de Complementaridad No Lineal), definido de la siguiente manera:

Hallar un $x^* \in \mathbb{R}_+^Q$ tal que $F(x^*) \cdot x^* = 0$, donde

$$\begin{aligned} F(x) &= [Y(x), Z(x)] \quad \text{con} \quad x = (h, \mu) \\ Y_r(x) &= C_r(h) - \mu_i \qquad \forall r \in \mathcal{P}_i, i \in \mathcal{I} \\ Z_i(x) &= \sum_{r \in \mathcal{P}_i} h_r - H_i(\mu) \qquad \forall i \in \mathcal{I} \\ Q &= P + I \qquad (\text{Notar que } x \in \mathbb{R}_+^Q = \mathbb{R}_+^{P+I}) \end{aligned} \tag{4.18}$$

En esta formulación, las primeras P componentes de F se refieren a los caminos y las restantes I, a los pares O/D.

Finalmente, para transformar este problema en una Desigualdad Variacional, aplicaremos el siguiente teorema:

Teorema 4.24. El punto $x^* \in \mathbb{R}_+^Q$ es una solución para el problema de complementaridad (4.18)

$$\begin{aligned} F(x^*) \cdot x^* &= 0 \\ F(x^*) &\geq 0 \end{aligned}$$

si, y solamente si x^* es una solución de la Desigualdad Variacional

$$F(x^*) \cdot (x - x^*) \geq 0 \qquad \forall x \in \mathbb{R}_+^Q$$

Capítulo 5

Conclusiones

Con este trabajo podemos concluir que:

1.- Se puede generalizar la teoría de optimización para problemas en los que la función de estudio es no diferenciable, en cuyo caso se trabajaría con la teoría del subdiferencial.

2.- Las Desigualdades Variacionales resumen problemas complejos en una sola desigualdad.

Bibliografía

Cohn-1980] COHN, D.L.: "Measure Theory", Birkhauser (Boston), 1980

Rudin-1973] WALTER RUDIN: "Functional Analysis", McGraw-Hill,Inc, 1973

Rockafellar-1970] R. TYRRELL ROCKAFELLAR: "Convex Analysis", Princeton University Press, 1970

Makela-1990] M. MÄKELÄ: "Nonsmooth optimization. Theory and algorithms with applications to optimal control", 1990

EuJa-2000] EUSEBI JARAUTA BRAGULAT: "Análisis matemático de una variable", Edicions UPC (Barcelona), 2000

Aashtiani-Magnanti-1981] H.Z. AASHTIANI AND T.L. MAGNANTI: "Equilibria on a congested transportation network. SIAM Journal on Algebraic and Discrete Methods 2", 1981