Capítulo 1

Optimización con restricciones

La optimización con restricciones se refiere a la minimización de funciones sujeta a restricciones sobre las variables. Una formulación general de estos problemas es

$$\min_{x \in \Re^n} f(x) \text{ sujeto a } \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I}, \end{cases} \tag{1.1}$$

donde las funciones f y c_i son funciones suaves de valor real definidas sobre un subconjunto de $\Re^n$, y $\mathcal{E}$ e $\mathcal{I}$ son dos conjuntos finitos de indices. En lo sucesivo se llamará a f la función objetivo, mientras que c_i, $i \in \mathcal{E}$ y c_i, $i \in \mathcal{I}$ son las *restricciones* de *igualdad* y de *desigualdad* respectivamente. Se define el conjunto factible Ω como el conjunto de puntos que satisfacen las restricciones, esto es,

$$\Omega = \{x \mid c_i(x) = 0,\ i \in \mathcal{E};\ \ c_i(x) \geq 0,\ i \in \mathcal{I}\}, \tag{1.2}$$

así se puede reescribir el problema (1.1) de forma más compacta como

$$\min_{x \in \Omega} f(x). \tag{1.3}$$

En este capítulo se derivarán las condiciones que caracterizan las soluciones de (1.3).

1.1 Soluciones locales y globales

Las definiciones de los diferentes tipos de soluciones locales son extensiones sencillas de las definiciones correspondientes para el caso sin restricciones, excepto que ahora nos limitamos a considerar puntos factibles en la vecindad de x^*. Se tienen las siguientes definiciones.

Un vector x^* es una *solución local* del problema (1.3) si $x^* \in \Omega$ y existe una vecindad $\mathcal{N}$ de x^* tal que $f(x) \geq f(x^*)$ para $x \in \mathcal{N} \cap \Omega$.

Un vector x^* es una *solución local estricta* (también *llamada solucion local fuerte*) si $x^* \in \Omega$ y existe una vecindad $\mathcal{N}$ de x^* tal que $f(x) > f(x^*)$ para $x \in \mathcal{N} \cap \Omega$ con $x \neq x^*$.

Ejemplo 1.1.1

$$\min x_1 + x_2 \ \ \textit{sujeto a} \ \ \ x_1^2 + x_2^2 - 2 = 0. \tag{1.4}$$

En este caso se tiene $f(x) = x_1 + x_2$, $\mathcal{I} = \emptyset$, $\mathcal{E} = \{1\}$, y $c_1(x) = x_1^2 + x_2^2 - 2$. Por inspección se puede ver que el conjunto factible es el circulo de radio $\sqrt{2}$ centrado en el origen (únicamente la frontera del circulo no su interior). No es difícil verificar que la solución x^* es $(-1, -1)^T$. Desde cualquier otro punto sobre el

Figure 1.1: Problema (1.4), se muestran los gradientes de las restricciones y de la función objetivo en varios puntos factibles.

circulo es sencillo encontrar un camino en el que los puntos que lo formen sean factibles y al mismo tiempo se este decrementando la funcion a lo largo de el. Por ejemplo, partiendo del punto $x = (\sqrt{2}, 0)^T$ cualquier movimiento en la dirección del movimiento de las manecillas del reloj alrededor del circulo tiene el efecto deseado.

De la Figura (1.1) podemos ver que en la solución x^*, la *restricción normal* $\nabla c_1(x^*)$ es paralela a $\nabla f(x^*)$. Esto es, existe un escalar λ_1^* tal que

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*). \tag{1.5}$$

(En este caso se tiene $\lambda_1^* = -\frac{1}{2}$.)

Se puede encontrar analíticamente (1.5) al aproximar la función objetivo y las restricciones con su serie de Taylor de primer orden correspondiente. Para mantener la factibilidad con respecto a la funcion $c_1(x) = 0$, se requiere que $c_1(x + d) = 0$; esto es,

$$0 = c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d. \tag{1.6}$$

De donde, la dirección d se mantiene factible con respecto a c_1, cuando esta satisface

$$\nabla c_1(x)^T d = 0. \tag{1.7}$$

De forma similar, una dirección de mejora debe generar un decremento en f, así que

$$0 > f(x + d) - f(x) \approx \nabla f(x)^T d,$$

o bien,

$$\nabla f(x)^T d < 0. \tag{1.8}$$

Si existe una dirección d que satisface (1.7) y (1.8) entonces se concluye que es posible tener una mejora en el punto actual x. De aqui se sigue que una condición *necesaria* de optimalidad para el problema (1.4) es que no exista una dirección d que satisfaga tanto (1.7) como (1.8).

Por medio de un dibujo, el lector puede verificar que la única forma en que dicha dirección no exista es que $\nabla f(x)$ y $\nabla c_1(x)$ sean paralelos, esto es, $\nabla f(x) = \lambda_1 \nabla c_1(x)$ para algún escalar λ_1. En caso de que esta condición no se satisfaga, la dirección definida por

$$d = -\left(I - \frac{\nabla c_1(x) \nabla c_1(x)^T}{\|\nabla c_1(x)\|^2} \right) \nabla f(x) \tag{1.9}$$

satisface las condiciones (1.7) y (1.8).

Ahora se introduce la *función Lagrangiana*

$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x), \tag{1.10}$$

y observe que $\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x)$ de modo que se puede reescribir la condición (1.5) como: en la solución x^*, existe un escalar λ_1^* tal que

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0. \tag{1.11}$$

Esto último sugiere que la solución del problema con restricciones de igualdad se calcula encontrando los puntos estacionarios de la función de Lagrange. El escalar λ_1 es llamado el *multiplicador de Lagrange* para la restricción $c_1(x) = 0$.

Observar que la condición (1.5) (equivalentemente 1.11) parece ser una condición *necesaria* para una solución óptima de el problema (1.4), pero no es *suficiente*. Por ejemplo en el problema (1.4), (1.5) se satisface en el punto $x = (1, 1)$ (con $\lambda = \frac{1}{2}$) pero este punto no es una solución (de hecho este punto maximiza la función objetivo en el circulo). Más aún, en el caso de problemas con restricciones de igualdad, la condición (1.5) no puede considerarse suficiente simplemente colocando alguna restricción sobre el signo de λ_1. Para hacer más claro este comentario, considere reemplazar la restricción $x_1^2 + x_2^2 - 2 = 0$ por su negativo $2 - x_1^2 - x_2^2 = 0$ en el ejemplo (1.4). La solución del problema no se ve afectada pero el valor de λ_1^* que satisface la condicioón (1.5) cambia de $\lambda_1^* = -\frac{1}{2}$ a $\lambda_1^* = \frac{1}{2}$.

Ahora considerese el ejemplo anterior pero con una pequeña modificación, se reemplaza la restricción de igualdad por una de desigualdad como sigue

Ejemplo 1.1.2

$$\min x_1 + x_2 \quad \textit{sujeto a} \quad x_1^2 + x_2^2 - 2 \geq 0. \tag{1.12}$$

En este caso la región factible es el ciruclo del problema (1.4) y su interior. Por inspección se observa que la solución sigue siento $(-1, -1)$ y que la condición (1.5) se cumple para $\lambda_1^* = \frac{1}{2}$. Sin embargo se tiene que el multiplicador de Lagrange tiene el signo diferente al encontrado en el ejemplo anterior.

Como antes la conjetura es que un punto x no es optimo si es posible encontrar un paso d que manteniendo la factibilidad decremente la función objetivo f considerando la aproximación de primer orden de la función. Entonces la dirección d decrementa el valor de la función objetivo si $\nabla f(x)^T d < 0$, mientras que mantiene la factibilidad si

$$0 \leq c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d,$$

es decir,

$$c_1(x) + \nabla c_1(x)^T d \geq 0. \tag{1.13}$$

Para determinar cuando existe una dirección que satisfaga (1.8) y (1.13) se consideran dos casos.

Caso I: Considerar que x esta estrictamente dentro del circulo, de modo que la desigualdad estricta $c_1(x) > 0$ se satisface. En este caso cualquier vector d cumple (1.13), teniendo cuidado que su longitud sea suficientement pequeña. En particular, siempre que $\nabla f(x^*) \neq 0$, se puede obtener una dirección d que satisgace (1.8) y (1.13) al tomar

$$d = -c_1(x) \frac{\nabla f(x)}{\|\nabla f(x)\|}.$$

La única situación en la que esta dirección no existe es cuando

$$\nabla f(x) = 0. \tag{1.14}$$

Caso II: Ahora considérese el caso en que x esta en la frontera del circulo, de modo que $c_1(x) = 0$. Las condiciones (1.8) y (1.13) son entonces

$$\nabla f(x)d \leq 0, \;\; \nabla c_1(x)^T d \geq 0. \tag{1.15}$$

Figure 1.2: Una dirección d que satisface (1.8) y (1.13) que vive en la intersección de un medio plano cerrado y un medio plano abierto.

La primera condición define un espacio abierto mientras que la segunda define un espacio cerrado (ver Figura 1.2). De la figura es claro que las dos regiones no se intersectan unicamente cuando $\nabla f(x)$ y $\nabla c_1(x)$ apuntan en la misma dirección, esto es, cuando

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \;\; \text{para algún } \lambda_1 \geq 0. \tag{1.16}$$

Observar que el signo del multiplicador es relevante. Si (1.5) hubiera sido satisfecha con un valor *negativo* de λ_1, entonces $\nabla f(x)$ y $\nabla c_1(x)$ apuntarían en dirección opuesta y entonces las direcciones que satisfacen (1.8) y (1.13) formarían un medio plano abierto.

Las condicione óptimas en ambos casos pueden ser resumidas como sigue. Cuando no existen direcciones factibles de descenso de primer orden en algún punto x^*, se tiene que

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0, \;\; \text{para algún } \lambda_1^* \geq 0, \tag{1.17}$$

donde también se requiere que

$$\lambda_1^* c_1(x^*) = 0. \tag{1.18}$$

Esta condición es conocida como la *condición de complementariedad*, ésta implica que el multiplicador de Lagrange λ_1 puede ser estrictamente positivo *únicamente cuando la restricción correspondiente c_1 esta activa*. En el caso I, se tenía que $c_1(x^*) > 0$, así que (1.18) requiere que $\lambda_1^* = 0$. Luego, (1.17) se reduce a $\nabla f(x^*) = 0$, como se necesitaba en (1.15). En el caso II, (1.18) permite a λ_1^* tomar valores no negativos, así (1.17) resulta ser equivalente a (1.16).

Ahora se agrega una restricción al problema 1.12

Ejemplo 1.1.3

$$\min x_1 + x_2 \;\; \textit{sujeto a} \;\;\; x_1^2 + x_2^2 - 2 \geq 0, x_2 \geq 0, \tag{1.19}$$

para el cual la región factible es la mitad del disco que se muestra en el (Figura 1.3). No es difícil ver que la solución es $(-\sqrt{2}, 0)^T$, un punto en el que ambas restricciones estan activas. Al repetir los argumentos que se utilizaron en los ejemplos anteriores, se concluye que una dirección d es una dirección de descenso factible de primer orden, si satisface las siguientes condiciones:

$$\nabla c_i(x)^T \geq 0, \;\; i \in \mathcal{I} = 1, 2, \;\; \nabla f(x)^T d < 0. \tag{1.20}$$

Sin embargo es claro de la (Figura 1.3) que esta dirección no puede existir cuando $x = (-\sqrt{2}, 0)^T$. Las condiciones $\nabla c_i(x)^T \geq 0, \; i = 1, 2$ son satisfechas ambas solo si d esta en el cuadrante definido por $\nabla c_1(x)$ y $\nabla c_2(x)$, pero por inspección es claro que todos los vectores d en este cuadrante satisfacen $\nabla f(x)^T d \geq 0$.

Figure 1.3: Una dirección d que satisface (1.8) y (1.13) que vive en la intersección de un medio plano cerrado y un medio plano abierto.

Ahora se revisa el comportamiento de el Lagrangiano y sus derivadas en el problema (1.19) y el punto solución $(-\sqrt{2}, 0)^T$. Primero se incluye un término adicional $\lambda_1 c_i(x)$ en el Lagrangiano para cada restricción adicional, de modo que se tiene

$$\mathcal{L}(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x),$$

donde $\lambda = (\lambda_1, \lambda_2)^T$ es el vector de multiplicadores de Lagrange. La extensión de la condicioón (1.17) en este caso es

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \;\; \text{para algún } \lambda^* \geq 0, \tag{1.21}$$

donde la desigualdad $\lambda^* \geq 0$ significa que todos los componentes de λ^* son no negativos. Al aplicar las condiciones de complementariedad (1.18) a ambas restricciones de desigualdad, se obtiene

$$\lambda_1^* c_1^*(x) = 0, \;\; \lambda_2^* c_2^*(x) = 0. \tag{1.22}$$

Cuando $x^* = (-\sqrt{2}, 0)^T$, tenemos

$$\nabla f(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{1.23}$$

así que es sencillo verificar que $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$ cuando se selecciona λ^* como sigue:

$$\lambda^* = \begin{bmatrix} 1/(2\sqrt{2}) \\ 1 \end{bmatrix}. \tag{1.24}$$

Observar que las componentes de λ^* son positivas.

Ahora se consideran algunos otros puntos factibles que no son solución de (1.19) y se examinan las propiedades de el Lagrangiano y su gradiente en estos puntos.

Para el punto $x = (\sqrt{2}, 0)^T$ se tiene que ambas restricciones estan activas. Sin embargo el gradiente de la función objetivo $\nabla f(x)$ ya no vive en el cuadrante definido por las condiciones $\nabla c_i(x)^T d \geq 0$, $i = 1, 2$. Una dirección de descenso factible de primer orden de este punto (un vector que satisface (1.20)) es simplemente $d = (-1, 0)^T$; hay muchos otros. Para este valor de x es sencillo verificar que la condición $\nabla_x \mathcal{L}(x, \lambda) = 0$ se satisface únicamente cuando $\lambda = (-1/(2\sqrt{2}), 1)$. Observar que la primera componente λ_1 es negativa, así que las condiciones (1.21) no se cumplen en este punto.

Finalmente, consideremos el punto $x = (1, 0)^T$, en el cual solo la segunda restricción c_2 esta activa. En este punto, linealizaciones de f y de c como en el Ejemplo 1.1.2 dan las siguientes condiciones, las cuales deben cumplirse para que d sea una dirección de descenso factible, de primer orden:

$$1 + \nabla c_1(x)^T d \geq 0, \quad \nabla c_2(x)^T d \geq 0, \quad \nabla f(x)^T d < 0. \tag{1.25}$$

De hecho solo se debe verificar que se satisfagan la segunda y tercera condición ya que siempre se satisface la primera al multiplicar d por una cantidad positiva lo suficientemente pequeña. Observar que

$$\nabla f(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla c_2(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{1.26}$$

es sencillo verificar que el vector $d = -(\frac{1}{2}, \frac{1}{4})$ satisface (1.25) y que es por lo tanto una dirección de descenso.

Para mostrar que las condiciones de optimalidad (1.21) y (1.22) fallan, se observar de (1.22) que debido a que $c_1(x) > 0$, se debe cumplir que $\lambda_1 = 0$. Por lo tanto, al tratar de satisfacer $\nabla_x \mathcal{L}(x, \lambda) = 0$, se esta buscando un valor λ_2 tal que $\nabla f(x) - \lambda_2 \nabla c_2(x) = 0$. Ya que λ_2 con esta característica no existe, este punto no satisface las condiciones de optimalidad.

1.2 Condiciones de optimalidad de primer orden

Establecimiento de las condiciones necesarias de primer orden

Ahora se generalizan las observaciones hechas en los ejemplos anteriores.

En general, el Lagrangiano para el problema de optimización con restricciones (1.1) se define como

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x). \tag{1.27}$$

El *conjunto activo* $\mathcal{A}(x)$ para cualquier punto factible x es la unión del conjunto $\mathcal{E}$ con los indices de las restricciones de desigualdad activas, esto es,

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\}. \tag{1.28}$$

Es importante considerar que es posible que $\nabla c_i(X)$ se anule debido a la representación algebraica de c_i, de modo que el término $\lambda_i \nabla c_i(x)$ se anula para todos los valores de λ_i. Lo que usualmente se hace para evitar este comportamiento degenerado en el valor de x en cuestion es establecer un supuesto sobre las restricciones como sigue.

Definición 1.2.1 *(LICQ). Dado el punto x^* y el conjunto activo $\mathcal{A}(x)$ definido por (1.28) se dice que se satisface la condición de independencia lineal en las restricciones (por sus siglas en inglés LICQ) si el conjunto de gradientes de las restricciones activas $\nabla c_i(x^*), i \in \mathcal{A}(x^*)$ es linealmente independiente.*

Esta condición permite establecer las condiciones de optimalidad para el problema general de optimización con restricciones (1.1).

Theorem 1.2.2 *(Condiciones de optimalidad de primer orden) Sea x^* una solución local del problema (1.1) en donde la condición LICQ se satisface, entonces hay un vector multiplicador de Lagrange λ^*, con componentes $\lambda_i^*, i \in \mathcal{E} \cup \mathcal{I}$, tal que las siguientes condiciones se satisfacen en (x^*, λ^*)*

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{1.29a}$$

$$c_i(x^*) = 0, \quad \textit{para toda } \; i \in \mathcal{E}, \tag{1.29b}$$

$$c_i(x^*) \geq 0, \quad \textit{para toda } \; i \in \mathcal{I}, \tag{1.29c}$$

$$\lambda_i^* \geq 0, \quad \textit{para toda } \; i \in \mathcal{I}, \tag{1.29d}$$

$$\lambda_i^* c_i(x^*) = 0, \quad \textit{para toda } \; i \in \mathcal{E} \cup \mathcal{I}. \tag{1.29e}$$

Las condiciones (1.29) son conocidas como las condiciones de *Karush-Kuhn-Tucker* o de forma más breve como las condiciones *KKT*. Debido a que la condición de complementariedad implica que los multiplicadores de Lagrange correspondientes a las restricciones de desigualdad son cero, es posible omitir los terminos de indices $i \notin \mathcal{A}(x^*)$ en (1.29a) y reescribir la condición como

$$0 = \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*) \tag{1.1}$$

Definición 1.2.3 *(Complementariedad estricta) Dada una solución local x^* de (1.1) y un vector λ^* que satisface (1.29), se dice que la condición de complementariedad estricta se satisface siempre que solo uno de los siguientes términos se anula, λ_i^* y $c_i(x^*)$ para cada índice $i \in \mathcal{I}$. En otras palabras, $\lambda_i^* > 0$ para toda $i \in \mathcal{I} \cup \mathcal{A}(x^*)$.*

Para un problema (1.1) y una solución x^* dados, existen muchos vectores λ^* para los que las condiciones (1.29) se cumplen. Sinembargo cuando la condición LICQ se satisface el óptimo λ^* es único.

Analísis de sensibilidad

El valor de cada multiplicador de Lagrange λ_i^* nos dice algo sobre la *sensibilidad* del valor de la función objetivo $f(x^*)$ a la presencia de la restricción c_i.

Considérese una restricción i activa, y perturbese el lado derecho de esta restricción un poco, digamos $c_i(x) \geq -\epsilon\|\nabla c_i(x^*)\|$ en lugar de $c_i(x) \geq 0$. Suponer que ϵ es lo suficientemente pequeña de modo que la solución perturbada $x^*(\epsilon)$ continúa teniendo el mismo conjunto de restricciones activas y que los multiplicadores de Lagrange no fueron afectados demasiado con la pertubación. Se tiene entonces que

$$\begin{aligned} -\epsilon\|\nabla c_i(x^*)\| &= c_i(x^*(\epsilon)) - c_i(x^*) \approx (x^*(\epsilon) - x^*)^T \nabla c_i(x^*), \\ 0 &= c_j(x^*(\epsilon)) - c_j(x^*) \approx (x^*(\epsilon) - x^*)^T \nabla c_j(x^*), \qquad (1.2) \\ & \qquad \text{para toda } \; j \in \mathcal{A}(x^*) \text{ con } j \neq i \qquad (1.3) \end{aligned}$$

El valor de $f(x^*(\epsilon))$, puede ser estimado con ayuda de (1.29a). Se tiene

$$\begin{aligned} f(x^*(\epsilon)) - f(x^*) &\approx (x^*(\epsilon) - x^*)^T \nabla f(x^*) \\ &= \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* (x^*(\epsilon) - x^*)^T \nabla c_j(x^*) \qquad (1.4) \\ &\approx -\epsilon\|\nabla c_i(x^*)\| \lambda_i^*. \qquad (1.5) \end{aligned}$$

Al tomar el límite se tiene que la familia de soluciones $x^*(\epsilon)$ satisface

$$\frac{df(x^*(\epsilon))}{d\epsilon} = -\lambda_i^* \|\nabla c_i(x^*)\|. \qquad (1.6)$$

Un análisis de sensibilidad de este problema podría concluir que si $\lambda_i^*\|\nabla c_i(x^*)\|$ es grande, entonces el valor óptimo es sensible a la existencia de la i-ésima restricción, mientras que si es pequenõ la dependencia no es tan fuerte. Si λ_i^* es exactamente cero para alguna restricción activa una pequenã perturbación a c_i en algunas direcciones dificilmente afectará el valor de la función objetivo.

Es importante hacer notar que el análisis anterior vale aún cuando las restricciones sean escaladas.

1.3 Derivación de las condiciones de primer orden

Sucesiones factibles

Dado un punto factible x^*, una sucesión $\{z_k\}_{k=0}^{\infty}$ con $z_k \in \Re^n$ es una *sucesión factible* si se satisfacen las siguientes propiedades:

i. $z_k \neq x^*$ para toda k;

ii. $lim_{k\to\infty} z_k = x^*$;

iii. z_k es factible para todo valor suficientemente grande de k.

Se define el conjunto de todas las sucesiones factibles que se acercan a x por $\mathcal{T}(x)$.

Se caracteriza una solución local de (1.1) como un punto x en el cual todas las sucesiones factibles tienen la propiedad de que $f(z_k) \geq f(x)$ para toda k suficientemente grande. En lo que sigue se derivan condiciones que se pueden verificar bajo las cuales esta propiedad se cumple.

Las *direcciones límite* de una sucesión factible son vectores d tales que

$$\lim_{z_k \in S_d} \frac{z_k - x}{\|z_k - x\|} \to d, \tag{1.7}$$

donde S_d es alguna subsucesión de $\{z_k\}_{k=0}^{\infty}$. En general, una sucesión factible tiene almenos una dirección factible y muchas tienen más de una.

Dado un punto x, si es posible elegir una sucesión factible de $\mathcal{T}(x)$ tal que la aproximación de primer orden de la función objetivo se incremente monótonamente a lo largo de la sucesión , entonces x no puede ser un punto óptimo.

Theorem 1.3.1 *Sea x^* una solución local de (1.1), entonces todas las sucesiones factibles $\{z_k\}$ en $\mathcal{T}(x^*)$ deben satisfacer*

$$\nabla f(x^*)^T d \geq 0, \tag{1.8}$$

donde d es alguna dirección límite de la sucesión factible.

Caracterizando las direcciones límite

El Teorema 1.3.1 es muy general sin embargo no es sencillo utilizarlo pues al parecer se hace necesario conocer todas las direcciones límite de todas las sucesiones factibles en $\mathcal{T}(x)$. En esta sección se muestra que ciertas cualidades en las restricciones permitirán caracterizar las propiedades más importantes de $\mathcal{T}(x)$, y por lo tanto se podrá verificar la condición (1.37) de modo más sencillo.

En los resultados siguientes se introduce la notación A para representar la matriz cuyas filas son los gradientes de las restricciones activas en el punto óptimo, esto es,

$$\nabla c_i^* = \nabla c_i(x^*), \quad A^T = [\nabla c_i^*]_{i\in\mathcal{A}(x^*)}, \quad \nabla f^* = \nabla f(x^*), \tag{1.9}$$

Lema 1.3.2 *Las siguientes dos afirmaciones son verdaderas*

i. Siempre que $d \in \Re^n$ sea una dirección límite de una sucesión factible, entonces

$$d^T \nabla c_i^* = 0, \;\; \textit{para toda} \;\; i \in \mathcal{E}, \;\; d^T \nabla c_i^* \geq 0, \;\; \textit{para toda} \;\; i \in \mathcal{A}(x^*) \cap \mathcal{I}. \tag{1.10}$$

ii. Si (1.39) se cumple con $\|d\| = 1$ y la condición LICQ se satisface también, entonces $d \in \Re^n$ es una dirección límite de alguna sucesión factible.

El conjunto de direcciones definido por (1.39) tiene un papel importante en las condiciones de optimalidad, así que para referencias futuras se le da un nombre al conjunto y se define formalmente.

Definición 1.3.3 *Dado un punto x^* y el conjunto de restricciones activas $\mathcal{A}(x^*)$ definido por (1.28), el conjunto F_1 se define como*

$$F_1 = \left\{ \alpha d | \alpha > 0, \begin{array}{ll} d^T \nabla c_i^* = 0, & \textit{para toda } \ i \in \mathcal{E} \\ d^T \nabla c_i^* \geq 0, & \textit{para toda } \ i \in \mathcal{A}(x^*) \cap \mathcal{I}, \end{array} \right\} \tag{1.11}$$

Observar que F_1 es un cono. De hecho, cuando una restricción es satisfecha, F_1 es el cono tangente al conjunto factible en x^*.

Introducción a los multiplicadores de Lagrange

El Lema (1.3.2) asegura que cuando la condición LICQ se cumple, el cono F_1 es simplemente el conjunto de todos los multiplos positivos de todas las direcciones límite de todas las sucesiones facibles posibles. Por lo tanto la condición (1.37) del Teorema 1.3.1 se cumple si $\nabla f(x^*)^T d < 0$ para toda $d \in F_1$. Esta condición también parece imposible de verificar ya que el conjunto F_1 tiene una infinidad de vectores. El siguiente lema brinda una alternativa.

Lema 1.3.4 *No existe dirección $d \in F_1$ para la cual $d^T \nabla f^* < 0$ si y solo si existe un vector $\lambda \in \Re^m$ con*

$$\nabla f^* = \sum_{i \in \mathcal{A}(x)} \lambda_i \nabla c_i^* = A(x^*)^T \lambda, \quad \lambda_i \geq 0 \ \textit{ para } \ i \in \mathcal{A}(x^*) \cap \mathcal{I}. \tag{1.12}$$

Prueba del Teorema 1.2.2

Los lemas (1.3.2) y (1.3.4) pueden combinarse para generar las condiciones KKT descritas en el Teorema 1.2.2. Suponer que $x \in \Re^n$ es un punto factible en el cual satisface la condición LICQ. El teorema asegura que si x^* es una solución local de (1.1), entonces existe un vector $\lambda^* \in \Re^m$ que satisface las condiciones (1.29).

En principio se muestra que hay multiplicadores $\lambda_i, i \in \mathcal{A}(x^*)$, tales que (1.41) se satisface. El Teorema 1.3.1 dice que $d^T \nabla f^* \geq 0$ para todos los vectores d tales que son direcciones límite de sucesiones factibles. A partir del Lema 1.3.2 se sabe que cuando las condiciones LICQ se cumplen, el conjunto de todas las direcciones limite posibles es exactamente el conjunto de vectores que satisfacen las condiciones) 1.39). Al colocar estas dos afirmaciones juntas, se encuentra que todas las direcciones d que satisfacen (1.39) también deben cumplir $d^T \nabla f^* \geq 0$. Luego, del Lema 1.3.4, se tiene que hay un vector λ para el cual (1.41) se vale, como se aseguró.

Ahora se define el vector λ^* como

$$\lambda_i^* = \left\{ \begin{array}{ll} \lambda_i, & i \in \mathcal{A}(x^*) \\ 0, & \text{en otro caso} \end{array} \right. \tag{1.13}$$

y se muestra que esta elección de λ^*, junto con la solución local x^*, satisface las condiciones (1.29). Se verifican estas condiciones una a una.

- La condición (1.29a) se sigue inmediatamente de (1.41) y de las definiciones (1.27) de la función Lagrangiano y de (1.42) de λ^*.
- Ya que x^* es factible, las condiciones (1.29b) y (1.29c) son satisfechas.

- Se tiene de (1.41) que $\lambda_i^* \geq 0$ para $i \in \mathcal{A}(x^*) \cap \mathcal{I}$, mientras que a partir de (1.42), $\lambda_i^* = 0$ para $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$. De donde, $\lambda_i^* \geq 0$ para $i \in \mathcal{I}$ así que (1.29d) se cumple.
- Se tiene para $i \in \mathcal{A}(x^*) \cap \mathcal{I}$ que $c_i(x^*) = 0$, mientras que para $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$ se tiene $\lambda_i^* = 0$. De donde $\lambda_i^* c_i(x^*) = 0$ para $i \in \mathcal{I}$ así que (1.29e) se satisface también.

1.4 Condiciones de segundo orden

Para las direcciones $w \in F_1$ para las cuales $w^T \nabla f(x^*) = 0$, no es posible determinar unicamente con la información de la primera derivada cuando un movimiento en esta dirección incrementará o decrementará la función objetivo f. Las condiciones de segundo orden examinan los términos de las segundas derivadas en la expansión por series de Taylor de f y de c_i, para ver cuando esta información extra resuelve la duda sobre el incremento o decremento de f.

Para el propósito de esta sección, se asume que f y c_i, $i \in \mathcal{E} \cup \mathcal{I}$, son dos veces continuamente diferenciables.

Dado F_1 de la definición (1.3.3) y algún vector de multiplicadores de Lagrange λ^* que satisface las condiciones KKT (1.29), se define un subconjunto $F_2(\lambda^*)$ de F_1 por

$$F_2(\lambda^*) = \{w \in F_1 | \nabla c_i(x^*)^T w = 0, \text{ toda } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ con } \lambda_i^* > 0\}.$$

Equivalentemente,

$$w \in F_2(\lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0, & \text{para toda } i \in \mathcal{E}, \\ \nabla c_i(x^*)^T w = 0, & \text{para toda } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ con } \lambda_i^* > 0, \\ \nabla c_i(x^*)^T w \geq 0, & \text{para toda } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ con } \lambda_i^* = 0. \end{cases}$$

A partir de esta última definición y del hecho de que $\lambda_i^* = 0$ para todo $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$, se sigue inmediatamente que

$$w \in F_2(\lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0 \text{ para toda } i \in \mathcal{E} \cup \mathcal{I}. \tag{1.14}$$

Luego, de la primera condición de KKT (1.29a) y de la definición (1.27) de la función Lagrangiano, se tiene que

$$w \in F_2(\lambda^*) \Rightarrow w^T \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* w^T \nabla c_i(x^*) = 0. \tag{1.15}$$

Theorem 1.4.1 *(Condiciones necesarias de segundo orden).*

Suponer que x^* es una solución local de (1.1) y que las condiciones LICQ se satisfacen. Sea λ^* un vector de multiplicadores de Lagrange tal que las condiciones KKT (1.29) se cumplen, y sea $F_2(\lambda^*)$ definido como arriba. Entonces

$$w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) \geq 0, \text{ para toda } w \in F_2(\lambda^*). \tag{1.16}$$

Theorem 1.4.2 *(Condiciones suficientes de segundo orden).*

Suponer que para algún punto factible $x^* \in \Re^n$ existe un vector de multiplicadores de Lagrange λ^* para el cual las condiciones de KKT (1.29) se cumplen. Suponer además que

$$w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \text{ para toda } w \in F_2(\lambda^*), w \neq 0. \tag{1.17}$$

Entonces x^* es una solución local estricta de (1.1).

Ejemplo

Ahora se vuelve a considerar el Ejemplo 1.1.2 para verificar las condiciones de segundo orden para el problema 1.12. En este problema se tenia $f(x) = x_1 + x_2$, $c_1(x) = 2 - x_1^2 - x_2^2, \mathcal{E} = \emptyset$, y $\mathcal{I} = \{1\}$. El Lagrangiano es

$$\mathcal{L}(x, \lambda) = (x_1 + x_2) - \lambda_1(2 - x_1^2 - x_2^2),$$

y es sencillo mostrar que las condiciones KKT (1.29) se cumplen por $x^* = (-1, -1)^T$, con $\lambda_1^* = \frac{1}{2}$. La Hessiana del Lagrangiano en este punto es

$$\nabla_{xx}\mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 2\lambda_i^*, & 0 \\ 0, & 2\lambda_i^* \end{bmatrix} = \begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix}. \tag{1.18}$$

Esta matriz es definida positiva, esto es, satisface que $w^T \nabla_{xx}\mathcal{L}(x^*, \lambda^*)w > 0$ para toda $w \neq 0$, así satisface las condiciones del Teorema 1.4.2. Se concluye que $x^* = (-1, -1)^T$ es una solución local estricta de (1.12). (De hecho, es la solución global de este problema.)

Referencias

[1] Studden, W. J. (1968). Optimal Designs on Tchebycheff Points. *The Annals of Mathematical Statistics*. Vol 39, **5**, 1435-1447.