Capítulo 3

Optimización con restricciones de desigualdad. Enfoque analítico.

3.1. Condiciones de Karush-Khun-Tucker

El problema a resolver es el siguiente: Determinar $\mathbf{x} \in \Re^n$ que minimiza la función $f : \Re^n \to \Re$ verificando el vector $\mathbf{x}$ un conjunto de restricciones $g_i(\mathbf{x}) \leq 0 \quad \forall i \in \{1, 2, \ldots, m\}$.

Al conjunto de soluciones del problema se le identifica por S:

$$S = \{\mathbf{x} \in \Re^n \ , \ g_i(\mathbf{x}) \leq 0 \ \forall i \in \{1, 2, \ldots, m\}\}$$

Conviene observar que todas las inecuaciones tendrán la forma anterior, es decir, desigualdades del tipo $\leq$ y dejando el término independiente a 0; esta restricción no supone pérdida de generalidad del problema a tratar.

Introduciendo m parámetros, que a diferencia del problema con restricciones de igualdad, denominaremos u_i con $i \in \{1, 2, \ldots, m\}$, uno por cada restricción, se define el lagrangiano:

$$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x})$$

A diferencia del capítulo anterior, no es suficiente con determinar los puntos estacionarios del lagrangiano para obtener las condiciones necesarias de mínimo; en su lugar, estas condiciones necesarias las aportan las condiciones de Karush-Khun-Tucker.

En el caso en que las funciones f y g_i tengan las condiciones analíticas adecuadas (diferenciabilidad, por ejemplo), las condiciones necesarias de mínimo para un vector $(\overline{\mathbf{x}}, \overline{\mathbf{u}}) \in \Re^{n+m}$, son:

1. Condición de gradiente, análoga a la de los multiplicadores de Lagrange:

$$\nabla L(\overline{\mathbf{x}}, \overline{u}) = \nabla f(\overline{\mathbf{x}}) + \overline{u}^t \nabla g(\overline{\mathbf{x}}) = \mathbf{0}$$

2. Condición de ortogonalidad, específica de las inecuaciones:

$$\overline{u}_i g_i(\overline{\mathbf{x}}) = 0 \qquad \forall i = 1, 2, \ldots, m$$

3. Condición de factibilidad, análoga a la de los multiplicadores de Lagrange:

$$g(\overline{\mathbf{x}}) \leq \mathbf{0}$$

4. Condición de no negatividad:

$$\overline{\mathbf{u}} \geq \mathbf{0}$$

Un vector $(\overline{\mathbf{x}}, \overline{\mathbf{u}}) \in \Re^{n+m}$ verificando estas condiciones se dice que es un punto KKT.

Estas condiciones permiten plantear el problema analíticamente (n+m ecuaciones y n+m incógnitas) distinguiendo ordenadamente los casos inducidos por la condición de ortogonalidad:

- $u_i = 0$, fijando una incógnita.
- $u_i > 0$, activando la ecuación $g_i(\mathbf{x}) = 0$.

Al igual que las condiciones de Lagrange, son condiciones necesarias de mínimo. La unicidad requiere que tanto la función f como las g_i sean convexas.

Para fijar ideas del procedimiento inducido por las condiciones KKT, se introduce el siguiente ejemplo.

Ejemplo 3.1. *Sea el problema en el plano ($n = 2$) con $m = 4$ inecuaciones:*

$$\begin{aligned} Min \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\ sujeto\ a \quad & x_1^2 + x_2^2 - 5 \leq 0 \\ & x_1 + 2x_2 - 4 \leq 0 \\ & -x_1 \leq 0 \\ & -x_2 \leq 0 \end{aligned}$$

Gráficamente, se tiene el recinto de soluciones factibles en la Figura 3.1.

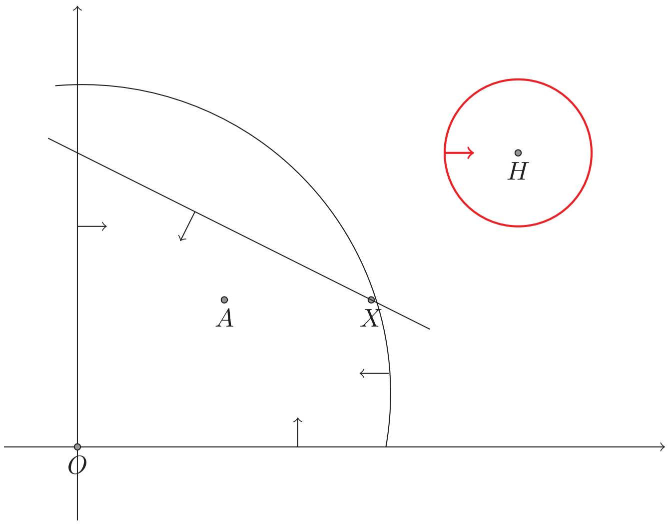

Figura 3.1: Representación gráfica del ejemplo 3.1

La disyuntiva entre el valor positivo o nulo de los u_i, asociada a la condición de ortogonalidad, permite el siguiente análisis:

- Un valor positivo $u_i > 0$ induce, por la condición de ortogonalidad, que $g_i(\mathbf{x}) = 0$ y, por tanto, $\mathbf{x}$ pertenece al borde de la inecuación.

- Si $g_i(\mathbf{x}) < 0$, es decir, $\mathbf{x}$ pertenece al interior del recinto i-ésimo, entonces el valor del multiplicador debe ser nulo ($u_i = 0$).

El desarrollo sistemático de la disyuntiva anterior permite realizar un análisis completo de los posibles óptimos locales de un problema de optimización con restricciones de desigualdad.

Otra propiedad interesante es la siguiente: si en un punto de la región factible $\mathbf{x} \in S$ el cono de direcciones de descenso no tiene intersección con el cono de direcciones factibles, será un punto KKT. A continuación se detalla esta idea.

Dado $\mathbf{x} \in S \subset \Re^n$, una solución factible del problema de optimización, una dirección es un vector $\mathbf{d} \in \Re^n$ que mantiene el punto dentro del recinto S con una magnitud de salto suficientemente pequeña; concretamente, existe un escalar $\lambda > 0$ suficientemente pequeño verificando:

$$\mathbf{x} + \lambda \mathbf{d} \in S$$

Dado $\mathbf{x} \in S$, una dirección de descenso es una dirección con la propiedad siguiente:

$$f(\mathbf{x} + \lambda \mathbf{d}) < f(\mathbf{x})$$

para el valor $\lambda > 0$ suficientemente pequeño que garantiza la pertenencia a S.

El cono de direcciones de descenso asociado a $\mathbf{x}$, denotado por $D(\mathbf{x})$, es el conjunto de sus direcciones de descenso; si se verifican las condiciones analíticas apropiadas, se puede caracterizar por:

$$D(\mathbf{x}) = \left\{ \mathbf{d} \in \Re^n \ , \ \nabla f(\mathbf{x})^t \mathbf{d} < 0 \right\}$$

Al conjunto de direcciones facibles de una solución factible $\mathbf{x} \in S$ se denomina cono de direcciones factibles y se denota por $F(\mathbf{x})$:

$$F(\mathbf{x}) = \{ \mathbf{d} \in \Re^n \ , \ \mathbf{x} + \lambda \mathbf{d} \in S \ , \ con \ \lambda > 0 \}$$

La caracterización analítica del conjunto de direcciones factibles es algo más compleja, pues depende del conjunto de restricciones activas asociadas a la solución $\mathbf{x} \in S$: una restricción $g_i(\mathbf{x})$ es activa si $g_i(\mathbf{x}) = 0$; en caso contrario, $g_i(\mathbf{x}) < 0$, se dice inactiva. Al conjunto de restricciones activas de $\mathbf{x}$ se denota por $I(\mathbf{x})$.

La relevancia de este concepto radica en que las restricciones inactivas no restringen el conjunto de soluciones factibles, pues para un salto $\lambda > 0$ suficientemente pequeño, mantienen la solución dentro de dicha restricción. No es el caso de las restricciones activas, pues al estar en el borde del recinto, restringen los movimientos que mantienen la solución dentro del mismo.

De hecho, la caracterización analítica del cono de direcciones factibles depende sólo de las restricciones activas:

$$F(\mathbf{x}) = \left\{\mathbf{d} \in \Re^n \ , \ \nabla g_i(\mathbf{x})^t \mathbf{d} < 0 \ \forall i \in I(\mathbf{x})\right\}$$

Continuando con el ejemplo 3.1, al ser $n = 2$, los distintos conceptos introducidos anteriormente pueden ser concretados y visualizados en diferentes puntos de la Figura 3.1:

- *Punto $H = (3, 2) \notin S$.*

  Es el óptimo global de la función f, pero no es una solución factible.

  Permite indicar la dirección de descenso máximo desde cualquier punto de $\Re^2$.

- *Punto $O = (0, 0) \in S$.*

  El conjunto de restricciones activas es $I(O) = \{3, 4\}$.

  $$\nabla f(O) = \begin{pmatrix} -6 \\ -4 \end{pmatrix} \qquad \nabla g_3(O) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \qquad \nabla g_4(O) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

  Los conjuntos de direcciones de descenso y factibles son:

  $$D(O) = \{(d_1, d_2) \in \Re^2 \ , \ 6d_1 + 4d_2 > 0\} \qquad F(O) = \{(d_1, d_2) \in \Re^2 \ , \ d_1, d_2 > 0\}$$

  La dirección $(1, 1) \in D(O) \bigcap F(O)$, O no puede ser un óptimo local.

- *Punto $A = (1, 1) \in S$.*

  El conjunto de restricciones activas es $I(A) = \emptyset$, es un punto interior del recinto.

  $$\nabla f(A) = \begin{pmatrix} -4 \\ -2 \end{pmatrix}$$

  Los conjuntos de direcciones de descenso y factibles son:

$$D(A) = \{(d_1, d_2) \in \Re^2 \ , \ 4d_1 + 2d_2 > 0\} \qquad F(A) = \{(d_1, d_2) \in \Re^2\}$$

La dirección $(1,1) \in D(A) \bigcap F(A)$*, A no puede ser un óptimo local.*

- *Punto* $X = (2,1) \in S$*, es el óptimo del problema.*

  El conjunto de restricciones activas es $I(X) = \{1,2\}$*.*

$$\nabla f(X) = \begin{pmatrix} -2 \\ -2 \end{pmatrix} \qquad \nabla g_1(X) = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \qquad \nabla g_2(X) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

  Los conjuntos de direcciones de descenso y factibles son:

$$D(X) = \{(d_1, d_2) \in \Re^2 \ , \ 2d_1 + 2d_2 > 0\}$$
$$F(X) = \{(d_1, d_2) \in \Re^2 \ , \ 4d_1 + 2d_2 < 0 \ , \ 1d_1 + 2d_2 < 0\}$$

  En este caso no existe una dirección que sea simultáneamente de descenso y factible, es decir,$D(X) \bigcap F(X) = \emptyset$ *y el punto X es un punto KKT, de hecho, es el óptimo local y global.*

Observación 3.1. *Si el problema incorpora restricciones en forma de ecuaciones e inecuaciones, se tratará aplicando a las inecuaciones el análisis KKT visto anteriormente e incorporando las ecuaciones del lagrangiano vistas en el capítulo anterior.*

3.2. Convexidad de la función objetivo y el recinto

Hasta ahora, tanto en este capítulo como en el anterior, los algoritmos de optimización (minimización) han tratado de buscar puntos estacionarios, en los que la derivada (dimensión 1) o gradiente (dimensión mayor que 1) se anulaba.

Estos puntos estacionarios identifican óptimos locales, y en el caso en que la función en dichos puntos sea convexa (segunda derivada positiva para dimensión 1 y matriz del hessiano definida positiva para dimensiones mayores) los óptimos serán mínimos locales.

En el caso de optimización sin restricciones, si la función es convexa en todo el espacio, el (único) mínimo local será un mínimo global. Esto es válido en el caso de optimización con restricciones de igualdad. En el caso de optimización con restricciones de desigualdad, sin embargo, el mínimo local será único y coincidirá con el mínimo global en el caso en que no sólo la función objetivo sea convexa sino que hay que exigir que el recinto de $\Re^n$ sea un conjunto convexo; más exactamente, para que la solución sea única en el punto $\mathbf{x}$ es suficiente con que sean convexas todas las funciones $g_i(\mathbf{x})$ para todo $i \in \{1, 2, \ldots, m\}$.

Para ilustrar este problema, se introduce otro ejemplo.

Ejemplo 3.2. *Sea el problema en el plano* $(n = 2)$ *con* $m = 3$ *inecuaciones:*

$$\begin{array}{rr} Min & 2x_1 - x_2 \\ sujeto\ a & (x_1 - 1)^2 + x_2 - 5 \leq 0 \\ & -x_1^2 + x_2 \leq 0 \\ & -x_2 \leq 0 \end{array}$$

Gráficamente, se tiene el recinto de soluciones factibles en la Figura 3.2.

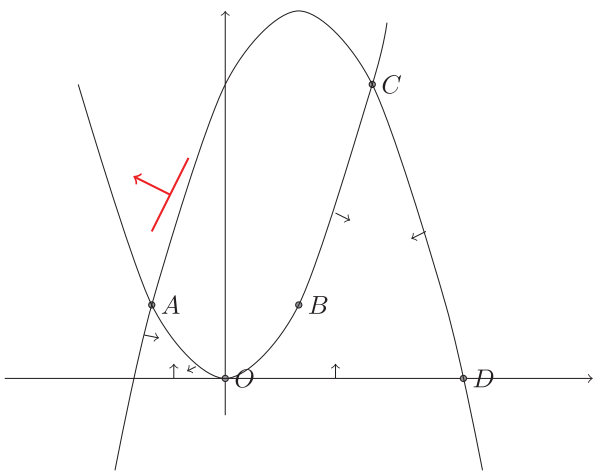

Figura 3.2: Representación gráfica del ejemplo 3.2

Se observa que la función $g_2(x_1, x_2) = -x_1^2 + x_2$ *no es una función convexa y el recinto* $\{(x_1, x_2) \in \Re^2\ ,\ g_2(x_1, x_2) \leq 0\}$ *no es un conjunto convexo.*

En este problema no hay un único punto KKT; en realidad hay tres puntos: $A = (-1, 1)$, $B = (1, 1)$ *y* $C = (2, 4)$.

3.3. El problema dual lagrangiano

Sea el problema de optimización que incluye restricciones de igualdad y desigualdad:

$$P : \begin{cases} Min & f(\mathbf{x}) \\ sujeto\ a: & g(\mathbf{x}) \leq \mathbf{0} \\ & h(\mathbf{x}) = \mathbf{0} \end{cases}$$

Se define el lagrangiano:

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^t g(\mathbf{x}) + \mathbf{v}^t h(\mathbf{x})$$

Se define el problema dual lagrangiano como:

$$D : \left\{ \max_{\mathbf{u} \geq \mathbf{0}, \mathbf{v}} \Theta(\mathbf{u}, \mathbf{v}) \right\}$$

siendo:

$$\Theta(\mathbf{u}, \mathbf{v}) \equiv Inf_{\{\mathbf{x}\}} \{L(\mathbf{x}, \mathbf{u}, \mathbf{v})\}$$

Los problemas P y D se denominan duales, y sus funciones objetivo verifican la desigualdad siguiente (teorema de dualidad débil):

$$\Theta(\mathbf{u}, \mathbf{v}) = Inf_{\{\mathbf{y}\}} \{L(\mathbf{y}, \mathbf{u}, \mathbf{v})\} \leq f(\mathbf{x}) + \mathbf{u}^t g(\mathbf{x}) + \mathbf{v}^t h(\mathbf{x}) \leq f(\mathbf{x})$$

para todo $\mathbf{x}$, solución factible de P ($g(\mathbf{x}) \leq \mathbf{0}, h(\mathbf{x}) = \mathbf{0}$) y $(\mathbf{u}, \mathbf{v})$ solución factible de D, es decir, $\mathbf{u} \geq \mathbf{0}$.

Como corolario, los óptimos de ambos problemas están acotados en el mismo sentido y, bajo ciertas condiciones de regularidad, alcanzan el mismo valor; esto ocurre, por ejemplo, en el caso lineal. Las condiciones de regularidad las impone el teorema fuerte de dualidad, que exige que las funciones f y g sean convexas, que la función h sea afín ($h(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$) , y exista un $\mathbf{x}^*$ verificando $g(\mathbf{x}^*) < \mathbf{0}$ y $h(\mathbf{x}^*) = \mathbf{0}$, en este caso:

$$Sup_{\{\mathbf{u} \geq \mathbf{0}, \mathbf{v}\}} \Theta(\mathbf{u}, \mathbf{v}) = Inf_{\{(\mathbf{x}, \mathbf{v}), g(\mathbf{x}) \leq \mathbf{0}, h(\mathbf{x}) = \mathbf{0}\}} \left\{ f(\mathbf{x}) + \mathbf{u}^t g(\mathbf{x}) + \mathbf{v}^t h(\mathbf{x}) \right\}$$

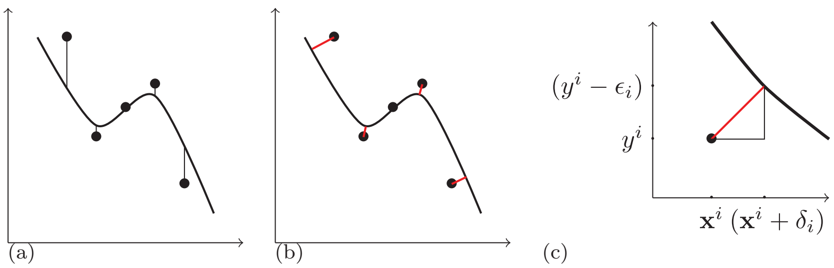

Figura 3.3: (a) Regresión no lineal; (b) Regresión ortogonal; y (c),detalle de los errores horizontal y vertical

Además, si este ínfimo es finito, se alcanza en un valor $\mathbf{x}^*$ y el supremo también se alcanza en el par $(\mathbf{u}^*, \mathbf{v}^*)$ con $\mathbf{u}^* \geq \mathbf{0}$ verificando la condición (de ortogonalidad) $\mathbf{u}^* g(\mathbf{x}^*) = 0$.

3.4. El modelo de regresión ortogonal

El modelo de regresión, tanto lineal como no lineal, minimiza el error cuadrático medio $\sum_i \epsilon_i^2$, siendo $\epsilon_i = (y^i - h(\mathbf{x}^i, \beta))$ para cada uno del m datos; gráficamente, ver Figura 3.3 (a), cada uno de estos ϵ_i mide la discrepancia *vertical* entre los puntos $(\mathbf{x}^i, y^i)$ y $(\mathbf{x}^i, h(\mathbf{x}^i, \beta)))$.

Existe otra posibilidad de ajustar la curva $h(\mathbf{x}^i, \beta))$ al conjunto de datos $\{(\mathbf{x}^i, y^i),\ i \in \{1, 2, \ldots, m\}\}$: sería aquélla que minimiza la distancia ortogonal entre la curva y la nube de puntos; es la situación descrita en la Figura 3.3 (b), donde se han señalado en rojo las discrepancias entre los citados puntos y la curva a ajustar mediante la elección adecuada de los parámetros β del modelo.

Se introduce así la **Regresión ortogonal**, que se detalla a continuación, pues se basa en un modelo de optimización no lineal con restricciones que generaliza el modelo de regresión no lineal. Para más detalles, ver Nocedal [20].

El error asociado a cada punto de la nube sería el siguiente:

$$y^i = h(\mathbf{x}^i + \delta^i, \beta)) + \epsilon_i \qquad \forall i \in \{1, 2, \ldots, m\}$$

Cada dato i-ésimo, ver Figura 3.3 (c), tiene asociados dos errores: ϵ_i, el error *vertical* que es análogo al error del modelo de regresión habitual, y δ_i, el error *horizontal* asociado al error cometido en las variables independientes $\mathbf{x}^i$ y que tiene dimensión n; en el caso descrito en la figura citada la dimensión de δ es 1; la distancia de la curva al punto (en rojo) es la distancia euclídea.

Observación 3.2. *En la representación de la Figura 3.3 (c), el valor* $\epsilon_i < 0$.

La suma de estas distancias euclídeas es la función a minimizar en el modelo de regresión ortogonal, que consiste en determinar el parámetro $\widehat{\beta}$ que minimiza la distancia euclídea de la nube de puntos a la curva $h(\mathbf{x}, \beta)$:

$$S(\widehat{\beta}) = \min_{\beta}\{S(\beta)\} = \min_{\beta}\left\{\sum_{i=1}^{m}(\epsilon_i^2 + \delta_i^2)\right\}$$

con las restricciones siguientes:

$$y^i = h(\mathbf{x}^i + \delta^i, \beta)) + \epsilon_i \qquad \forall i \in \{1, 2, \ldots, m\}$$