Concavidad y Optimización

Abelardo Jordán Liza

Maestría en Matemáticas Aplicadas
PUCP

Setiembre 26, 2023

Ahora considere el problema

$$(P) \qquad \begin{array}{ll} \text{mín} & f(x) \\ s.a. & g_i(x) \leq 0,\ i = 1, \cdots, m. \\ & x \in C \end{array} \tag{1}$$

donde $f, g_i, i = 1, \cdots, m$ son funciones diferenciables en $\mathbb{R}^n$ y C es un subconjunto convexo no vacío de $\mathbb{R}^n$. Si x^* resuelve (P), entonces existen escalares $\lambda_0, \lambda_1, \cdots, \lambda_m$ no negativos y no todos cero, tales que

$$\begin{array}{l} (i) 0 \in \ \lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x*) + N_C(x^*) \\ (ii)\ \lambda_i g_i(x^*) = 0,\ i = 1, \cdots, m. \end{array} \tag{2}$$

Ahora considere el problema

$$(P) \qquad \begin{array}{ll} \text{mín} & f(x) \\ s.a. & g_i(x) \leq 0,\ i = 1, \cdots, m. \\ & x \in C \end{array} \tag{1}$$

donde $f, g_i, i = 1, \cdots, m$ son funciones diferenciables en $\mathbb{R}^n$ y C es un subconjunto convexo no vacío de $\mathbb{R}^n$. Si x^* resuelve (P), entonces existen escalares $\lambda_0, \lambda_1, \cdots, \lambda_m$ no negativos y no todos cero, tales que

$$\begin{array}{l} (i) 0 \in \ \lambda_0 \nabla f(x^*) + \displaystyle\sum_{i=1}^{m} \lambda_i \nabla g_i(x*) + N_C(x^*) \\ (ii)\ \lambda_i g_i(x^*) = 0,\ i = 1, \cdots, m. \end{array} \tag{2}$$

Dem: Análogo a la demostración de la proposición anterior, probando que el sistema

$$(I) : \langle \nabla f(x^*), v \rangle < 0 \ \ ; \ \ g_i(x^*) + \langle \nabla g_i(x^*), v \rangle < 0,\ i = 1, \cdots, m \tag{3}$$

no tiene solución para $v \in T_C(x^*)$.

Ahora considere el problema

$$(P) \qquad \begin{array}{rl} \text{mín} & f(x) \\ s.a. & g_i(x) \leq 0,\ i = 1, \cdots, m. \\ & x \in C \end{array} \tag{1}$$

donde $f, g_i, i = 1, \cdots, m$ son funciones diferenciables en $\mathbb{R}^n$ y C es un subconjunto convexo no vacío de $\mathbb{R}^n$. Si x^* resuelve (P), entonces existen escalares $\lambda_0, \lambda_1, \cdots, \lambda_m$ no negativos y no todos cero, tales que

$$\begin{array}{l} (i) 0 \in \ \lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x*) + N_C(x^*) \\ (ii)\ \lambda_i g_i(x^*) = 0,\ i = 1, \cdots, m. \end{array} \tag{2}$$

Dem: Análogo a la demostración de la proposición anterior, probando que el sistema

$$(I) : \langle \nabla f(x^*), v \rangle < 0 \ \ ; \ \ g_i(x^*) + \langle \nabla g_i(x^*), v \rangle < 0,\ i = 1, \cdots, m \tag{3}$$

no tiene solución para $v \in T_C(x^*)$.Aplicando el teorema de la alternativa, existen $\lambda_0, \lambda_1, \cdots, \lambda_m$ no negativos y no todos cero, tales que

$$\lambda_0\langle\nabla f(x^*),v\rangle+\sum_{i=1}^{m}\lambda_i(g_i(x^*)+\langle\nabla g_i(x^*),v\rangle)\geq 0,\ \forall v\in T_C(x^*) \tag{4}$$

lo que indica que

$$-(\lambda_0\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*))\in(T_C(x^*))^\circ$$

Para el problema (P), denotemos por $I(x^*)$ al conjunto $\{i \in \{1, \cdots, m\} : g_i(x^*) = 0\}$ llamado el conjunto de los índices de las restricciones activas en x^*.

Corolario

Si en el problema (P), $x^* \in int(C)$ y los vectores $\{\nabla g_i(x^*)\}_{i \in I(x^*)}$ son l.i. Entonces existen $\lambda_1, \cdots, \lambda_m \geq 0$ tales que

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*) = 0$$

$$\lambda_i g_i(x^*) = 0, \ i = 1, \cdots, m.$$

Sean C un subconjunto convexo de $\mathbb{R}^n$, $f : C \to \mathbb{R}$ una función diferenciable, $A : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal, $b \in \mathbb{R}^m$.Considere el problema

$$(P=) \qquad \begin{array}{ll} \min & f(x) \\ s.a. & Ax = b \\ & x \in C \end{array} \tag{5}$$

Si $x^* \in int(C)$ resuelve (P=), entonces $\nabla f(x^*) \in A^t \mathbb{R}^m$.

Sean C un subconjunto convexo de $\mathbb{R}^n$, $f : C \to \mathbb{R}$ una función diferenciable, $A : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal, $b \in \mathbb{R}^m$.Considere el problema

$$(P =) \qquad \begin{array}{ll} \text{mín} & f(x) \\ s.a. & Ax = b \\ & x \in C \end{array} \tag{5}$$

Si $x^* \in int(C)$ resuelve (P=), entonces $\nabla f(x^*) \in A^t\mathbb{R}^m$.
Dem: Sea X el conjunto factible, entonces

$$\begin{aligned} y \in N_X(x^*) \quad & \Leftrightarrow \langle y, x - x^* \rangle \leq 0, \forall x : Ax = b, x \in C \\ & \Leftrightarrow \langle y, z \rangle \leq 0, \forall z : Az = 0. \\ & \Leftrightarrow \langle y, z \rangle = 0, \forall z \in Nu(A). \\ & y \in Im(A^t) \end{aligned}$$

lo que nos dice que $y \in A^t(\mathbb{R}^m)$, en consecuencia $\nabla f(x^*) \in A^t(\mathbb{R}^m)$.

Sean C un subconjunto convexo de $\mathbb{R}^n$, $f : C \to \mathbb{R}$ una función diferenciable, $A : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal, $b \in \mathbb{R}^m$.Considere el problema

$$(P=) \qquad \begin{array}{ll} \text{mín} & f(x) \\ s.a. & Ax = b \\ & x \in C \end{array} \tag{5}$$

Si $x^* \in int(C)$ resuelve (P=), entonces $\nabla f(x^*) \in A^t\mathbb{R}^m$.
Dem: Sea X el conjunto factible, entonces

$$\begin{aligned} y \in N_X(x^*) \quad & \Leftrightarrow \langle y, x - x^* \rangle \leq 0, \forall x : Ax = b, x \in C \\ & \Leftrightarrow \langle y, z \rangle \leq 0, \forall z : Az = 0. \\ & \Leftrightarrow \langle y, z \rangle = 0, \forall z \in Nu(A). \\ & y \in Im(A^t) \end{aligned}$$

lo que nos dice que $y \in A^t(\mathbb{R}^m)$, en consecuencia $\nabla f(x^*) \in A^t(\mathbb{R}^m)$. recuerde que la condición necesaria de optimalidad es $-\nabla f(x^*) \in N_X(x^*)$, pero como el último conjunto es un subespacio vectorial, equivale a $\nabla f(x^*) \in A^t(\mathbb{R}^m)$, esto significa que existe un vector $y \in \mathbb{R}^m$ tal que $\nabla f(x^*) = A^t y$.

Sean f, g_i, h_j funciones continuamente diferenciables en $\mathbb{R}^n$ y C un conjunto convexo cerrado. Para el problema

$$(ProMix) \qquad \begin{array}{ll} \text{mín} & f(x) \\ s.a. & g_i(x) \leq 0, \ i = 1, \cdots, m. \\ & h_j(x) = 0, \ j = 1, \cdots, p. \\ & x \in C \end{array} \tag{6}$$

sea X su conjunto factible.

Sean f, g_i, h_j funciones continuamente diferenciables en $\mathbb{R}^n$ y C un conjunto convexo cerrado. Para el problema

$$(ProMix) \qquad \begin{array}{ll} \text{mín} & f(x) \\ s.a. & g_i(x) \leq 0,\ i = 1, \cdots, m. \\ & h_j(x) = 0,\ j = 1, \cdots, p. \\ & x \in C \end{array} \tag{6}$$

sea X su conjunto factible.¿Cómo describir $N_X(x)$?

Considere el sistema

$$\begin{array}{l} g_i(x) \leq 0,\ i = 1, \cdots, m. \\ h_j(x) = 0,\ j = 1, \cdots, p. \\ x \in C \end{array}$$

cuyo conjunto solución X se asume no vacío. Se asume que f, g_i, h_j son funciones continuamente diferenciables en $\mathbb{R}^n$ y C es un conjunto convexo cerrado. Para un punto $\hat{x} \in X$, sea $I(\hat{x}) := \{i \in \{1, \cdots, m\} \,:\, g_i(\hat{x}) = 0\}$ (la colección de los sub-índices de las restricciones de desigualdad activas en $\hat{x}$)

Considere el sistema

$$\begin{array}{l} g_i(x) \leq 0,\ i = 1, \cdots, m. \\ h_j(x) = 0,\ j = 1, \cdots, p. \\ x \in C \end{array}$$

cuyo conjunto solución X se asume no vacío. Se asume que f, g_i, h_j son funciones continuamente diferenciables en $\mathbb{R}^n$ y C es un conjunto convexo cerrado. Para un punto $\hat{x} \in X$, sea $I(\hat{x}) := \{i \in \{1, \cdots, m\} : g_i(\hat{x}) = 0\}$ (la colección de los sub-índices de las restricciones de desigualdad activas en $\hat{x}$)
Si $G(x) = (g_1(x), \cdots, g_m(x); h_1(x), \cdots, h_p(x))'$ y
$Y_0 := (-\mathbb{R}^m_+) \times \{0\}_p \subset \mathbb{R}^m \times \mathbb{R}^p$, entonces el sistema puede escribirse como

$$\begin{array}{c} G(x) \in Y_0 \\ x \in C \end{array} \tag{7}$$

Considere el sistema

$$\begin{aligned} & g_i(x) \leq 0,\ i = 1, \cdots, m. \\ & h_j(x) = 0,\ j = 1, \cdots, p. \\ & x \in C \end{aligned}$$

cuyo conjunto solución X se asume no vacío. Se asume que f, g_i, h_j son funciones continuamente diferenciables en $\mathbb{R}^n$ y C es un conjunto convexo cerrado. Para un punto $\hat{x} \in X$, sea $I(\hat{x}) := \{i \in \{1, \cdots, m\} : g_i(\hat{x}) = 0\}$ (la colección de los sub-índices de las restricciones de desigualdad activas en $\hat{x}$)
Si $G(x) = (g_1(x), \cdots, g_m(x); h_1(x), \cdots, h_p(x))'$ y
$Y_0 := (-\mathbb{R}^m_+) \times \{0\}_p \subset \mathbb{R}^m \times \mathbb{R}^p$, entonces el sistema puede escribirse como

$$\begin{gathered} G(x) \in Y_0 \\ x \in C \end{gathered} \tag{7}$$

Definición

(Condición de Robinson) Sea $\hat{x} \in X$, se dice que (7)satisface la condición de Robinson en $\hat{x}$ si se cumple

$$\left\{G'(\hat{x})d - v \ : d \in K_C(\hat{x})\, , v \in K_{Y_0}(G(\hat{x}))\right\} = \mathbb{R}^{m+p} \tag{8}$$

- Note que el conjunto del lado izquierdo de (8) es un cono de $\mathbb{R}^{m+p}$ ¿Cuando un cono coincide con el espacio vectorial ambiente? Esto ocurre si 0 pertenece al interior del cono, es decir la condición de Robinson equivale a

$$0 \in int\left(\left\{G'(\hat{x})d - v \ : d \in K_C(\hat{x})\,,\, v \in K_{Y_0}(G(\hat{x}))\right\}\right)$$

- Note que el conjunto del lado izquierdo de (8) es un cono de $\mathbb{R}^{m+p}$ ¿Cuando un cono coincide con el espacio vectorial ambiente? Esto ocurre si 0 pertenece al interior del cono, es decir la condición de Robinson equivale a

$$0 \in int\left(\left\{G'(\hat{x})d - v \ : d \in K_C(\hat{x})\,,\, v \in K_{Y_0}(G(\hat{x}))\right\}\right)$$

- Si en (8), $C = \mathbb{R}^n$, entonces (8) toma la forma

$$\left\{G'(\hat{x})d - v \ : d \in \mathbb{R}^n\,,\, v \in K_{Y_0}(G(\hat{x}))\right\} = \mathbb{R}^{m+p}$$

- Note que el conjunto del lado izquierdo de (8) es un cono de $\mathbb{R}^{m+p}$ ¿Cuando un cono coincide con el espacio vectorial ambiente? Esto ocurre si 0 pertenece al interior del cono, es decir la condición de Robinson equivale a

$$0 \in int\left(\left\{G'(\hat{x})d - v \ : d \in K_C(\hat{x})\,,\, v \in K_{Y_0}(G(\hat{x}))\right\}\right)$$

- Si en (8), $C = \mathbb{R}^n$, entonces (8) toma la forma

$$\left\{G'(\hat{x})d - v \ : d \in \mathbb{R}^n\,,\, v \in K_{Y_0}(G(\hat{x}))\right\} = \mathbb{R}^{m+p}$$

- Si en (8), $C = \mathbb{R}^n$, $Y_0 = \{0\}$ entonces (8) se reduce a que

las filas de $G'(\hat{x})$ sean linealmente independientes .

Teorema

Si el sistema satisface la condición de Robinson en $\hat{x}$, entonces

$$T_X(\hat{x}) = \left\{d \in \mathbb{R}^n \ : d \in T_C(\hat{x})\ , G'(\hat{x})d \in T_{Y_0}(G(\hat{x}))\right\} \tag{9}$$

Teorema

Si el sistema satisface la condición de Robinson en $\hat{x}$, entonces

$$T_X(\hat{x}) = \left\{ d \in \mathbb{R}^n \ : d \in T_C(\hat{x})\ , G'(\hat{x})d \in T_{Y_0}(G(\hat{x})) \right\} \tag{9}$$

¿Existe alguna condición suficiente para la condición de Robinson?

Teorema

Si el sistema satisface la condición de Robinson en $\hat{x}$, entonces

$$T_X(\hat{x}) = \left\{d \in \mathbb{R}^n \ : d \in T_C(\hat{x}) \ , G'(\hat{x})d \in T_{Y_0}(G(\hat{x}))\right\} \tag{9}$$

¿Existe alguna condición suficiente para la condición de Robinson?

Proposición

(Mangasarian) Asuma que existe $x_M \in int(C)$ tal que

$$\begin{aligned} \langle \nabla g_i(\hat{x}), x_M - \hat{x}\rangle &< 0 \qquad , i \in I(\hat{x}) \\ \langle \nabla h_j(\hat{x}), x_M - \hat{x}\rangle &= 0 \qquad j = 1, \cdots, p. \end{aligned}$$

y los gradientes $\nabla h_j(\hat{x})$, $j = 1, \cdots, p$ son linealmente independientes, entonces el sistema satisface la condición de Robinson en $\hat{x}$ y concecuentemente se cumple el teorema previo.

Sean f, g_i, h_j funciones continuamente diferenciables en $\mathbb{R}^n$ y C un conjunto convexo cerrado. Para el problema

$$(ProMix) \qquad \begin{array}{ll} \text{mín} & f(x) \\ s.a. & g_i(x) \leq 0,\ i = 1, \cdots, m. \\ & h_j(x) = 0,\ j = 1, \cdots, p. \\ & x \in C \end{array} \tag{10}$$

Suponga que adicionalmente las funciones g_i son convexas y las funciones h_j son afines. Se dice que el sistema satisface la **Condición de Slater** si existe $x_s \in C$ tal que $g_i(x_s) < 0$ para cada $i = 1, \cdots, m$, $h_j(x_s) = 0$ para cada $j = 1, \cdots, p$; adicionalmente $x_s \in int(C)$ si $p > 0$.

Teorema

Asumiendo que la condición de Slater se satisface, entonces para $\hat{x} \in X$, se cumple

$$(T_X(\hat{x}))^{\circ} = (T_C(\hat{x}))^{\circ} + cone\left(\{\nabla g_i(\hat{x})\}_{i \in I(\hat{x})}\right) + span\{\nabla h_j(\hat{x}) : j = 1, \cdots, p\} \tag{11}$$

Teorema

Asumiendo que la condición de Slater se satisface, entonces para $\hat{x} \in X$, se cumple

$$(T_X(\hat{x}))^\circ = (T_C(\hat{x}))^\circ + cone\left(\{\nabla g_i(\hat{x})\}_{i\in I(\hat{x})}\right) + span\{\nabla h_j(\hat{x}) : j = 1, \cdots, p\} \tag{11}$$

Retomando el problema (ProMix), si las funciones f, g_i son convexas y las h_j son afines y C es convexo cerrado, entonces $\hat{x} \in X$ resuleve (ProMix) si y solo si

$$-\nabla f(\hat{x}) \in (T_X(\hat{x}))^\circ.$$

Si el sistema satisface la condición de Robinson o de Slater, se dice que el sistema satisface la **condición de calificación**.

Teorema

Sea $\hat{x} \in X$ un mínimo local del problema (ProMix) y asuma que en $\hat{x}$ se satisface una condición de calificación. Entonces existen $\lambda_i \geq 0$ para $i = 1, \cdots, m$ y $\mu_j \in \mathbb{R}$ para $j = 1, \cdots, p$ tales que

$$\begin{aligned} 0 \in \quad & \nabla f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) + \sum_{j=1}^{p} \mu_j \nabla h_j(\hat{x}) + N_C(\hat{x}) \\ & \lambda_i g_i(\hat{x}) = 0, \qquad i = 1, \cdots m. \end{aligned} \tag{12}$$

Si C es un conjunto convexo y cerrado, f, g_i son funciones convexas, h_j afines, entonces es válido el recíproco.

Si el sistema satisface la condición de Robinson o de Slater, se dice que el sistema satisface la **condición de calificación**.

Teorema

Sea $\hat{x} \in X$ un mínimo local del problema (ProMix) y asuma que en $\hat{x}$ se satisface una condición de calificación. Entonces existen $\lambda_i \geq 0$ para $i = 1, \cdots, m$ y $\mu_j \in \mathbb{R}$ para $j = 1, \cdots, p$ tales que

$$
\begin{aligned}
0 \in \quad & \nabla f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) + \sum_{j=1}^{p} \mu_j \nabla h_j(\hat{x}) + N_C(\hat{x}) \\
& \lambda_i g_i(\hat{x}) = 0, \qquad i = 1, \cdots m.
\end{aligned} \tag{12}
$$

Si C es un conjunto convexo y cerrado, f, g_i son funciones convexas, h_j afines, entonces es válido el recíproco.

El teorema es válido sin condición de Robinson si las funciones g_i, h_j son afines y C es un poliedro convexo.

Asociada al problema (ProMix), se introduce la *función Lagrangiana*, definida en $C \times \mathbb{R}_+^m \times \mathbb{R}^p$ por

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \tag{13}$$

Asociada al problema (ProMix), se introduce la *función Lagrangiana*, definida en $C \times \mathbb{R}^m_+ \times \mathbb{R}^p$ por

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \tag{13}$$

Por tanto, la condición necesaria para el (ProMix) puede expresarse como

$$-\nabla_x L(\hat{x}, \lambda, \mu) \in N_C(\hat{x}).$$

- (Minimizando una función de pérdida)Un modelo representa una variable respuesta $y \in \mathbb{R}$ como una función lineal de la forma

$$y = \sum_{i=1}^{n} x_i u_i$$

dependiente de n variables características $u_1, \cdots, u_n$. Las cantidadaes $x_1, \cdots, x_n$ son coeficientes desconocidos del modelo. Se tiene N observaciones $(u^j, y^j) \in \mathbb{R}^n \times \mathbb{R}$ para $j = 1, \cdots, N$.

- (Minimizando una función de pérdida)Un modelo representa una variable respuesta $y \in \mathbb{R}$ como una función lineal de la forma

$$y = \sum_{i=1}^{n} x_i u_i$$

dependiente de n variables características $u_1, \cdots, u_n$. Las cantidadaes $x_1, \cdots, x_n$ son coeficientes desconocidos del modelo. Se tiene N observaciones $(u^j, y^j) \in \mathbb{R}^n \times \mathbb{R}$ para $j = 1, \cdots, N$. Un criterio para elegir los coeficientes x_i es minimizar la suma de errores cuadrados:

$$f(x) = \sum_{j=1}^{N} \left(y^j - \sum_{i=1}^{n} x_i u_i^j \right)^2$$

- (Minimizando una función de pérdida)Un modelo representa una variable respuesta $y \in \mathbb{R}$ como una función lineal de la forma

$$y = \sum_{i=1}^{n} x_i u_i$$

dependiente de n variables características $u_1, \cdots, u_n$. Las cantidadaes $x_1, \cdots, x_n$ son coeficientes desconocidos del modelo. Se tiene N observaciones $(u^j, y^j) \in \mathbb{R}^n \times \mathbb{R}$ para $j = 1, \cdots, N$. Un criterio para elegir los coeficientes x_i es minimizar la suma de errores cuadrados:

$$f(x) = \sum_{j=1}^{N} \left(y^j - \sum_{i=1}^{n} x_i u_i^j \right)^2$$

Sean $Y = (y_1, \cdots, y_N)'$ y $U = [u_i^j]_{N \times n}$, entonces f toma la forma $f(x) = \|Y - Ux\|^2$ y se busca x que resuelva el problema

$$\begin{array}{cc} \text{mín} & \|Y - Ux\|^2 \\ x \in \mathbb{R}^n & \end{array}$$

- (Minimizando una función de pérdida)Un modelo representa una variable respuesta $y \in \mathbb{R}$ como una función lineal de la forma

$$y = \sum_{i=1}^{n} x_i u_i$$

dependiente de n variables características $u_1, \cdots, u_n$. Las cantidadaes $x_1, \cdots, x_n$ son coeficientes desconocidos del modelo. Se tiene N observaciones $(u^j, y^j) \in \mathbb{R}^n \times \mathbb{R}$ para $j = 1, \cdots, N$. Un criterio para elegir los coeficientes x_i es minimizar la suma de errores cuadrados:

$$f(x) = \sum_{j=1}^{N} \left(y^j - \sum_{i=1}^{n} x_i u_i^j \right)^2$$

Sean $Y = (y_1, \cdots, y_N)'$ y $U = [u_i^j]_{N\times n}$, entonces f toma la forma $f(x) = \|Y - Ux\|^2$ y se busca x que resuelva el problema

$$\min_{x \in \mathbb{R}^n} \quad \|Y - Ux\|^2$$

f es 2- diferenciable con $\nabla f(x) = 2U'(Ux - Y)$ y $Hf(x) = 2UU'$.

- (Minimizando una función de pérdida)Un modelo representa una variable respuesta $y \in \mathbb{R}$ como una función lineal de la forma

$$y = \sum_{i=1}^{n} x_i u_i$$

dependiente de n variables características $u_1, \cdots, u_n$. Las cantidadaes $x_1, \cdots, x_n$ son coeficientes desconocidos del modelo. Se tiene N observaciones $(u^j, y^j) \in \mathbb{R}^n \times \mathbb{R}$ para $j = 1, \cdots, N$. Un criterio para elegir los coeficientes x_i es minimizar la suma de errores cuadrados:

$$f(x) = \sum_{j=1}^{N} \left(y^j - \sum_{i=1}^{n} x_i u_i^j \right)^2$$

Sean $Y = (y_1, \cdots, y_N)'$ y $U = [u_i^j]_{N \times n}$, entonces f toma la forma $f(x) = \|Y - Ux\|^2$ y se busca x que resuelva el problema

$$\begin{array}{cc} \text{mín} & \|Y - Ux\|^2 \\ x \in \mathbb{R}^n & \end{array}$$

f es 2- diferenciable con $\nabla f(x) = 2U'(Ux - Y)$ y $Hf(x) = 2UU'$..
Note que $\nabla f(x) = 0 \Leftrightarrow U'Ux = UY$. Si U tiene rango n, entonces la solución es única $\hat{x} = (U'U)^{-1}U'Y$.

Previo: (Ejercicio): Sean $m \in \mathbb{N}$, $a^1, \cdots, a^m$ vectores de $\mathbb{R}^n$ y $\alpha_1, \cdots, \alpha_m$ números reales positivos. Se define la función H en $\mathbb{R}^n$ por

$$H(x) = \ln\left(\sum_{i=1}^{m} \alpha_i e^{\langle a^i, x\rangle}\right)$$

Probar que H es una función convexa.

Previo: (Ejercicio): Sean $m \in \mathbb{N}$, $a^1, \cdots, a^m$ vectores de $\mathbb{R}^n$ y $\alpha_1, \cdots, \alpha_m$ números reales positivos. Se define la función H en $\mathbb{R}^n$ por

$$H(x) = \ln \left(\sum_{i=1}^{m} \alpha_i e^{\langle a^i, x \rangle} \right)$$

Probar que H es una función convexa.

- Sea $\mathcal{M}$ una colección no vacía de subconjuntos de $\{1, 2, \cdots, n\}$ y $\theta \in \mathbb{R}^n$ dados. Para cada $M \in \mathcal{M}$, sea $\mathbb{1}_M$ el vector indicador del conjunto M. Se define en $\mathbb{R}^n$ la función

$$f(x) = \theta' x + \ln \left(\sum_{M \in \mathcal{M}} e^{\langle \mathbb{1}_M, x \rangle} \right)$$

y el problema de optimización global $\displaystyle \min_{x \in \mathbb{R}^n} \; f(x)$

Previo: (Ejercicio): Sean $m \in \mathbb{N}$, $a^1, \cdots, a^m$ vectores de $\mathbb{R}^n$ y $\alpha_1, \cdots, \alpha_m$ números reales positivos. Se define la función H en $\mathbb{R}^n$ por

$$H(x) = \ln\left(\sum_{i=1}^{m} \alpha_i e^{\langle a^i, x\rangle}\right)$$

Probar que H es una función convexa.

- Sea $\mathcal{M}$ una colección no vacía de subconjuntos de $\{1, 2, \cdots, n\}$ y $\theta \in \mathbb{R}^n$ dados. Para cada $M \in \mathcal{M}$, sea $\mathbb{1}_M$ el vector indicador del conjunto M. Se define en $\mathbb{R}^n$ la función

$$f(x) = \theta' x + \ln\left(\sum_{M\in\mathcal{M}} e^{\langle \mathbb{1}_M, x\rangle}\right)$$

y el problema de optimización global $\min_{x \in \mathbb{R}^n} \; f(x)$

Dado que f es convexa y continuamente diferenciable en $\mathbb{R}^{\ltimes}$, el problema tiene solución si la ecuación $\nabla f(x) = 0$ tiene solución x. Es decir

$$\theta + \frac{\sum\limits_{M\in\mathcal{M}} e^{\langle \mathbb{1}_M, x\rangle} \mathbb{1}_M}{\sum\limits_{M\in\mathcal{M}} e^{\langle \mathbb{1}_M, x\rangle}} = 0$$

Involucrando log-convex

Previo: (Ejercicio): Sean $m \in \mathbb{N}$, $a^1, \cdots, a^m$ vectores de $\mathbb{R}^n$ y $\alpha_1, \cdots, \alpha_m$ números reales positivos. Se define la función H en $\mathbb{R}^n$ por

$$H(x) = \ln\left(\sum_{i=1}^{m} \alpha_i e^{\langle a^i, x\rangle}\right)$$

Probar que H es una función convexa.

- Sea $\mathcal{M}$ una colección no vacía de subconjuntos de $\{1, 2, \cdots, n\}$ y $\theta \in \mathbb{R}^n$ dados. Para cada $M \in \mathcal{M}$, sea $\mathbb{1}_M$ el vector indicador del conjunto M. Se define en $\mathbb{R}^n$ la función

$$f(x) = \theta' x + \ln\left(\sum_{M \in \mathcal{M}} e^{\langle \mathbb{1}_M, x\rangle}\right)$$

y el problema de optimización global $\begin{array}{cc} \text{mín} & f(x) \\ x \in \mathbb{R}^n & \end{array}$

Dado que f es convexa y continuamente diferenciable en $\mathbb{R}^{\ltimes}$, el problema tiene solución si la ecuación $\nabla f(x) = 0$ tiene solución x. Es decir

$$\theta + \frac{\displaystyle\sum_{M \in \mathcal{M}} e^{\langle \mathbb{1}_M, x\rangle} \mathbb{1}_M}{\displaystyle\sum_{M \in \mathcal{M}} e^{\langle \mathbb{1}_M, x\rangle}} = 0$$

¿Cómo influye θ para esta ecuación tenga solución ?

- Sea A una matriz de orden $m \times n$ de rango m y $b \in \mathbb{R}^m$. Dado $z \in \mathbb{R}^n$ deseamos evaluar la distancia(Euclidiana) de z al conjunto afín $\{x \in \mathbb{R}^n : Ax = b\}$.

- Sea A una matriz de orden $m \times n$ de rango m y $b \in \mathbb{R}^m$. Dado $z \in \mathbb{R}^n$ deseamos evaluar la distancia(Euclidiana) de z al conjunto afín $\{x \in \mathbb{R}^n : Ax = b\}$.Entonces se genera el problema:

$$\begin{array}{ll} \min & \|z - x\| \\ s.a. & Ax = b \end{array}$$

el cual es equivalente a

$$\begin{array}{ll} \min & \|z - x\|^2 \\ s.a. & Ax = b \end{array}$$

- Sea A una matriz de orden $m \times n$ de rango m y $b \in \mathbb{R}^m$. Dado $z \in \mathbb{R}^n$ deseamos evaluar la distancia(Euclidiana) de z al conjunto afín $\{x \in \mathbb{R}^n : Ax = b\}$.Entonces se genera el problema:

$$\begin{array}{ll} \text{mín} & \|z - x\| \\ s.a. & Ax = b \end{array}$$

el cual es equivalente a

$$\begin{array}{ll} \text{mín} & \|z - x\|^2 \\ s.a. & Ax = b \end{array}$$

Se trata de un problema con m restricciones afines de igualdad, del tipo $h_j(x) = a_j x - b_j$, donde a_j es la $j-$ésima fila de A y $C = \mathbb{R}^n$. Por las condiciones de A automáticamente se cumple la condición de Robinson.

- Sea A una matriz de orden $m \times n$ de rango m y $b \in \mathbb{R}^m$. Dado $z \in \mathbb{R}^n$ deseamos evaluar la distancia(Euclidiana) de z al conjunto afín $\{x \in \mathbb{R}^n : Ax = b\}$.Entonces se genera el problema:

$$\begin{array}{ll} \text{mín} & \|z - x\| \\ s.a. & Ax = b \end{array}$$

el cual es equivalente a

$$\begin{array}{ll} \text{mín} & \|z - x\|^2 \\ s.a. & Ax = b \end{array}$$

Se trata de un problema con m restricciones afines de igualdad, del tipo $h_j(x) = a_j x - b_j$, donde a_j es la $j-$ésima fila de A y $C = \mathbb{R}^n$. Por las condiciones de A automáticamente se cumple la condición de Robinson. En tal caso, la función Lagrangiana es $L(x, \mu) = \|z - x\|^2 + \mu'(Ax - b)$ y la condición necesaria de optimalidad es

$$2(\hat{x} - z) + A'\mu = 0$$

¿Se conocen μ, $\hat{x}$?

- Sea A una matriz de orden $m \times n$ de rango m y $b \in \mathbb{R}^m$. Dado $z \in \mathbb{R}^n$ deseamos evaluar la distancia(Euclidiana) de z al conjunto afín $\{x \in \mathbb{R}^n : Ax = b\}$.Entonces se genera el problema:

$$\begin{array}{ll} \text{mín} & \|z - x\| \\ s.a. & Ax = b \end{array}$$

el cual es equivalente a

$$\begin{array}{ll} \text{mín} & \|z - x\|^2 \\ s.a. & Ax = b \end{array}$$

Se trata de un problema con m restricciones afines de igualdad, del tipo $h_j(x) = a_j x - b_j$, donde a_j es la $j-$ésima fila de A y $C = \mathbb{R}^n$. Por las condiciones de A automáticamente se cumple la condición de Robinson. En tal caso, la función Lagrangiana es $L(x,\mu) = \|z - x\|^2 + \mu'(Ax - b)$ y la condición necesaria de optimalidad es

$$2(\hat{x} - z) + A'\mu = 0$$

¿Se conocen μ, $\hat{x}$?En la última ecuación multiplicamos por A y así $\mu = 2(A'A)^{-1}(Az - b)$ y sustituyendo μ en la ecuación previa, podemos obtener $\hat{x} = z - \dfrac{1}{2}A'\mu = z - A'(AA')^{-1}(Az - b)$ (Verificar que $\hat{x}$ satisface $A\hat{x} = b$).

- Sea A una matriz de orden $m \times n$ de rango m y $b \in \mathbb{R}^m$. Dado $z \in \mathbb{R}^n$ deseamos evaluar la distancia(Euclidiana) de z al conjunto afín $\{x \in \mathbb{R}^n : Ax = b\}$.Entonces se genera el problema:

$$\begin{array}{ll} \text{mín} & \|z - x\| \\ s.a. & Ax = b \end{array}$$

el cual es equivalente a

$$\begin{array}{ll} \text{mín} & \|z - x\|^2 \\ s.a. & Ax = b \end{array}$$

Se trata de un problema con m restricciones afines de igualdad, del tipo $h_j(x) = a_j x - b_j$, donde a_j es la $j-$ésima fila de A y $C = \mathbb{R}^n$. Por las condiciones de A automáticamente se cumple la condición de Robinson. En tal caso, la función Lagrangiana es $L(x, \mu) = \|z - x\|^2 + \mu'(Ax - b)$ y la condición necesaria de optimalidad es

$$2(\hat{x} - z) + A'\mu = 0$$

¿Se conocen μ, $\hat{x}$?En la última ecuación multiplicamos por A y así $\mu = 2(A'A)^{-1}(Az - b)$ y sustituyendo μ en la ecuación previa, podemos obtener $\hat{x} = z - \dfrac{1}{2}A'\mu = z - A'(AA')^{-1}(Az - b)$ (Verificar que $\hat{x}$ satisface $A\hat{x} = b$).
El valor óptimo es $\|z - \hat{x}\| = \|A'(AA')^{-1}(Az - b)\|$

Caso particular: Distancia a un hiperplano

$$\begin{array}{ll} \text{mín} & \|z - x\| \\ s.a. & a'x = c \end{array}$$

$$\begin{array}{ll} \text{mín} & \|z-x\| \\ s.a. & a'x = c \end{array}$$

En este caso A es el vector fila no nulo a' de dimensión $1 \times n$.
$\hat{x}$ que resuelve el problema es $\hat{x} = a'(aa')^{-1}(az-c)$.
El valor óptimo es

$$\|z-\hat{x}\| = \frac{|a'z-c|}{\|a\|}$$