Optimización con Restricciones
Fundamentos y Marco Teórico

Dr. Ing. Claudio Risso (crisso@fing.edu.uy)

Instituto de Computación (FING - UDELAR)

Curso Optimización Continua y Aplicaciones (OCA)
(Octubre 2023)

Programación Matemática

El problema de referencia para esta parte del curso será el de programación matemática:

$$(P)\begin{cases} \min f(x) \\ g_i(x) \leq 0, \quad 1 \leq i \leq p, \\ h_j(x) = 0, \quad 1 \leq j \leq q, \\ x \in X \end{cases}$$

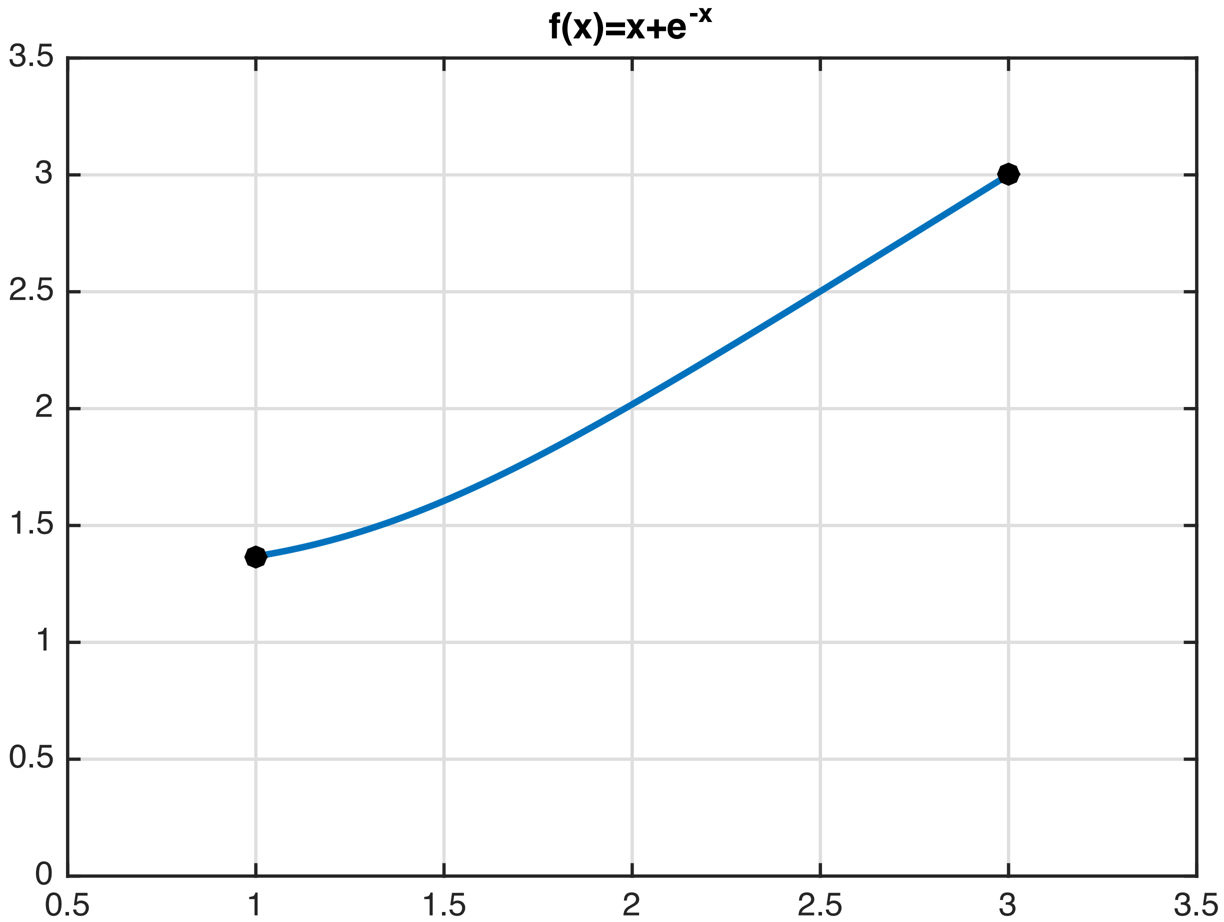
El problema (P) tiene n variables y $m = p + q$ restricciones.

Como norma, asumiremos que $X = \mathbb{R}^n$ si no se especifica.

Observar que incluso siendo todas las funciones diferenciables, ya no podemos asumir simplemente que un óptimo local del problema debe cumplir $\nabla f(x) = \vec{0}$.

Programación Matemática

Observar que incluso siendo todas las funciones diferenciables, ya no podemos asumir simplemente que un óptimo local del problema debe cumplir $\nabla f(x) = \vec{0}$.



Programación Matemática

El problema de referencia para esta parte del curso será el de programación matemática:

$$(P)\begin{cases} \min f(x) \\ g_i(x) \leq 0, \quad 1 \leq i \leq p, \\ h_j(x) = 0, \quad 1 \leq j \leq q, \\ x \in X \end{cases}$$

El problema (P) tiene n variables y $m = p + q$ restricciones.

Como norma, asumiremos que $X = \mathbb{R}^n$ si no se especifica.

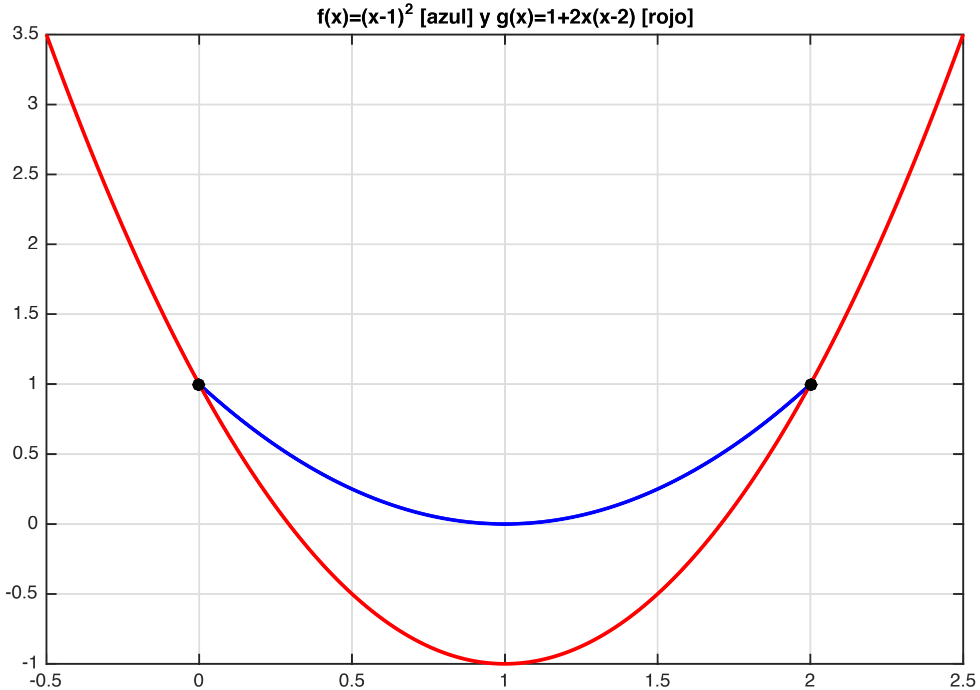
Observar que incluso siendo todas las funciones diferenciables, ya no podemos asumir simplemente que un óptimo local del problema debe cumplir $\nabla f(x) = \vec{0}$.

¿Qué propiedades teóricas debemos usar para encontrar óptimos en problemas con restricciones?

Optimización y Relajaciones

Definición

Dados dos problemas $(G)\begin{cases}\min f(x)\\ x \in F\end{cases}$ *y* $(G_L)\begin{cases}\min f_L(x)\\ x \in F_L\end{cases}$, *se dice que* (G_L) *es una relajación de* (G), *sii se cumplen las siguientes dos condiciones: i)* $F \subseteq F_L$ *y ii)* $f_L(x) \leq f(x), \forall x \in F$.



$$(P)\begin{cases}\min\,(x-1)^2\\ 0 \leq x \leq 2\end{cases}$$

$$(P_L)\begin{cases}\min 1 + 2x(x-2)\\ -\frac{1}{2} \leq x \leq \frac{5}{2}\end{cases}$$

(P_L) es una relajación de (P)

Optimización y Relajaciones

Propiedad

Sea (G_L) una relajación de (G), $\bar{x}_L$ y $\bar{x}$ óptimos de (G_L) y (G) respectivamente, y se cumple que $\bar{x}_L \in F$, entonces, también se cumple que: $f_L(\bar{x}_L) \leq f(\bar{x}) \leq f(\bar{x}_L)$.

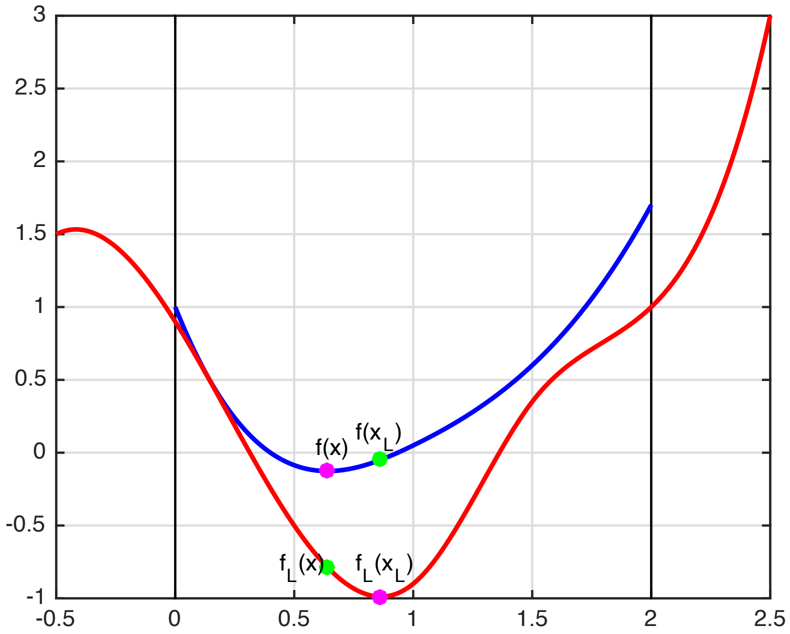
PRUEBA: Observar que $\bar{x} \in F_L$ (ya que $x \in F$ y $F \subseteq F_L$). Como $\bar{x}_L$ es el óptimo de (G_L), se cumple $f_L(\bar{x}_L) \leq f_L(\bar{x})$, ya que en (G_L), $\bar{x}$ es simplemente un punto más. Además, $f_L(\bar{x}) \leq f(\bar{x})$, porque $x \in F$ y $f_L(\cdot)$ es la función de la relajación. Eso concluye que $f_L(\bar{x}_L) \leq f(\bar{x})$, la primera parte de la desigualdad.

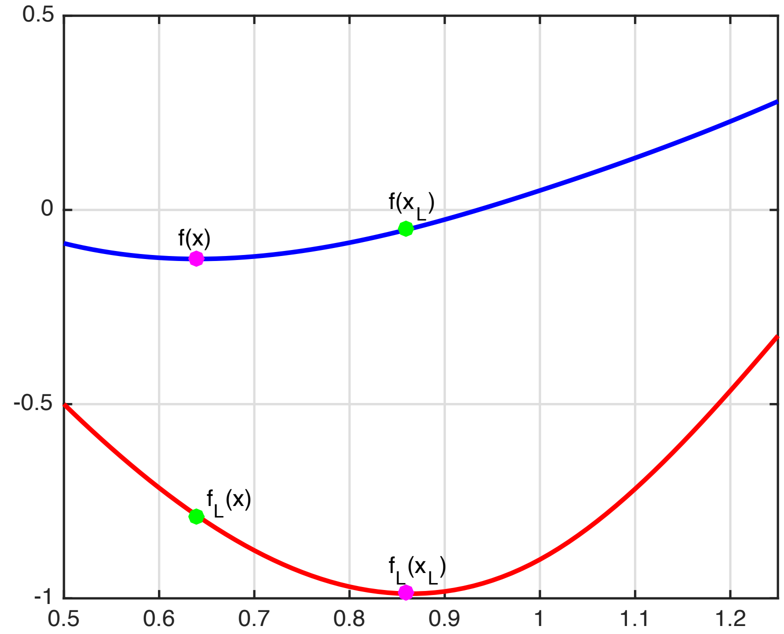
La segunda parte $(f(\bar{x}) \leq f(\bar{x}_L))$ se explica porque $\bar{x}$ es el óptimo de (G), $\bar{x}_L \in F$ (por hipótesis) y por tanto, ese punto es uno más en el problema del cual $\bar{x}$ es óptimo.

Optimización y Relajaciones

Propiedad

Sea (G_L) una relajación de (G), $\bar{x}_L$ y $\bar{x}$ óptimos de (G_L) y (G) respectivamente, y se cumple que $\bar{x}_L \in F$, entonces, también se cumple que: $f_L(\bar{x}_L) \leq f(\bar{x}) \leq f(\bar{x}_L)$.





Optimización y Relajaciones

Propiedad

Sea (G_L) una relajación de (G), $\bar{x}_L$ y $\bar{x}$ óptimos de (G_L) y (G) respectivamente, y se cumple que $\bar{x}_L \in F$, entonces, también se cumple que: $f_L(\bar{x}_L) \leq f(\bar{x}) \leq f(\bar{x}_L)$.

Corolario (Óptimo mediante Relajación)

Sea (G_L) una relajación de (G), y $\bar{x}_L$ la solución óptima de (G_L). Si se cumple que $\bar{x}_L \in F$ y $f_L(\bar{x}_L) = f(\bar{x}_L)$, entonces $\bar{x}_L$ también es solución óptima de (G).

Esta última propiedad es uno de los pilares para generalizar las condiciones teóricas a los problemas continuos con restricciones, e incluso las técnicas exactas en la optimización combinatoria.

Ejemplo (Relajación Entera)

Si las variables de un problema de optimización están definidas sobre el conjunto $X = \mathbb{Z}^n$, llamamos Relajación Entera al problema equivalente en el que $X_L = \mathbb{R}^n$.

Es una relajación porque los reales son más que los enteros, y cuando coinciden, la función objetivo vale lo mismo.

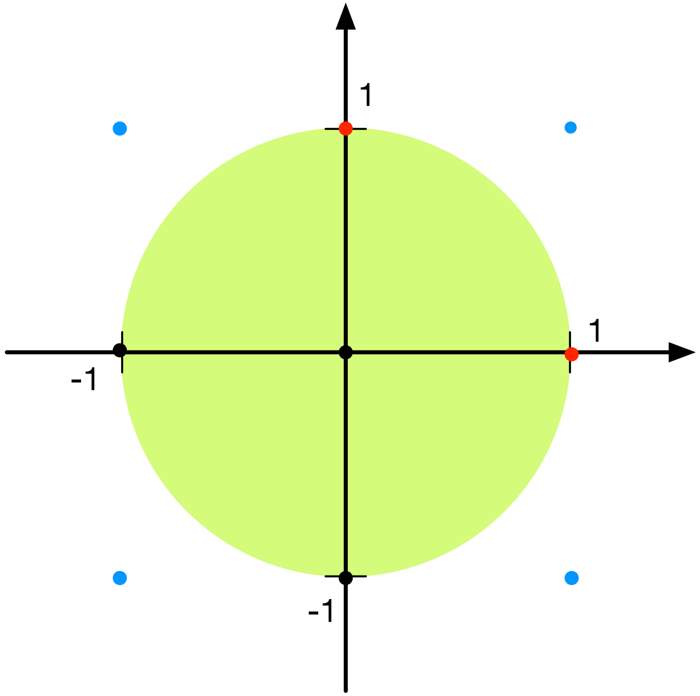
$$(P) \begin{cases} \min x^2 + 4y^2 - 12x - 8y + 10 \\ x, y \in \mathbb{Z} \end{cases}$$

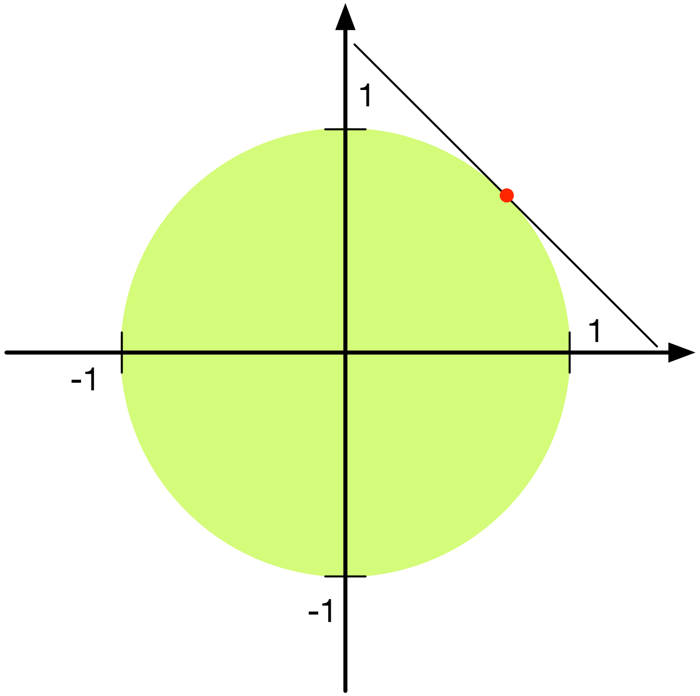
Ya lo resolvimos con (P_L) en $\mathbb{R}^2$ y la solución era (6,1), entonces ésa también es la solución para (P), que está definido en $\mathbb{Z}^2$.

El argumento no es general porque el óptimo en (P_L) puede no ser entero.

Ejemplo (Relajación Entera)

El argumento no es general porque el óptimo en (P_L) puede no ser entero.





Lo habíamos visto en el problema $(P)\left\{\begin{array}{l}\max x+y\\ x^2+y^2\leq 1\end{array}\right.$

Ejemplo (Relajación Entera)

El argumento no es general porque el óptimo en (P_L) puede no ser entero.

Lo que sí vale siempre es que el óptimo de (P_L) es una cota inferior del de (P).

Además, el mismo punto $\bar{x}$ solución de (P_L) tiene valor en sí mismo. Si por ejemplo, $\bar{x}_2 = 3.2$, podemos generar dos sub-problemas (proceso conocido como branching), uno en el que se agrega la restricción $x_2 \leq 3$ y otro en el que $x_2 \geq 4$.

Variantes de estas ideas se usan en los algoritmos *branch-and-boud*, *branch-and-cut*, etc., que son los Algoritmos Exactos por excelencia en la programación entera.

Esos algoritmos están fuera del alcance de este curso.

Relajación Lagrangeana

Es la reducción fundamental en la programación continua.

$$(P)\begin{cases} \min f(x) \\ g_i(x) \leq 0, \quad 1 \leq i \leq p, \\ h_j(x) = 0, \quad 1 \leq j \leq q \end{cases}$$

Dado un problema (P) y un conjunto de constantes $\lambda_i \geq 0$ $(1 \leq i \leq p)$ y μ_j $(1 \leq j \leq q)$, construye una relajación $(P_{\lambda,\mu})$ sin restricciones.

$$(P_{\lambda,\mu})\left\{ \min f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \mu_j h_j(x) \right.$$

Observar que $F \subseteq F_{\lambda,\mu}$, porque $(P_{\lambda,\mu})$ no tiene restricciones. Además, como en la región factible F de (P) se verifica $g_i(x) \leq 0$ y $h_j(x) = 0$, $\boldsymbol{f_{\lambda,\mu}(x)} = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \mu_j h_j(x) \leq f(x)$, ya que $\lambda_i g_i(x) \leq 0$ y $\mu_j h_j(x) = 0$.

Relajación Lagrangeana

Es la reducción fundamental en la programación continua.

$$(P)\begin{cases} \min f(x) \\ g_i(x) \leq 0, \quad 1 \leq i \leq p, \\ h_j(x) = 0, \quad 1 \leq j \leq q \end{cases}$$

Dado un problema (P) y un conjunto de constantes $\lambda_i \geq 0$ $(1 \leq i \leq p)$ y μ_j $(1 \leq j \leq q)$, construye una relajación $(P_{\lambda,\mu})$ sin restricciones.

$$(P_{\lambda,\mu})\left\{ \min f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \mu_j h_j(x) \right.$$

No es necesario relajar todas las restricciones para tener una Relajación Lagrangeana. Pero si lo hacemos, tendremos un problema donde sí vale que en el óptimo se debe anular el gradiente (cuando las funciones son diferenciables).

Relajación Lagrangeana

Es la reducción fundamental en la programación continua.

$$(P)\begin{cases} \min f(x) \\ g_i(x) \leq 0, \quad 1 \leq i \leq p, \\ h_j(x) = 0, \quad 1 \leq j \leq q \end{cases}$$

Dado un problema (P) y un conjunto de constantes $\lambda_i \geq 0$ $(1 \leq i \leq p)$ y μ_j $(1 \leq j \leq q)$, construye una relajación $(P_{\lambda,\mu})$ sin restricciones.

$$(P_{\lambda,\mu})\left\{ \min f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \mu_j h_j(x) \right.$$

Los λ's y μ's no son variables, son parámetros del problema relajado. Constantes, a efectos de derivar o usar propiedades.

Con eso en mente, se puede usar alguna de las técnicas de optimización sin restricciones vistas previamente en el curso.

Relajación Lagrangeana

Teorema (Fundamental de la Relajación Lagrangeana)

Sea $\bar{x}$ una solución de $(P_{\lambda,\mu})$ que cumple:

1. $g_i(\bar{x}) \leq 0 \, (1 \leq i \leq p)$ y $h_j(\bar{x}) = 0 \, (1 \leq j \leq q)$, *[$\bar{x}$ factible en (P)]*
2. $\lambda_i \geq 0 \, (1 \leq i \leq p)$, *[$(P_{\lambda,\mu})$ es una relajación de (P)]*
3. $\lambda_i \cdot g_i(\bar{x}) = 0 \, (1 \leq i \leq p)$, *[$f_{\lambda,\mu}(\bar{x}) = f(\bar{x})$]*

Entonces $\bar{x}$ también es solución óptima de (P).

PRUEBA: Se cumple porque se dan las condiciones del corolario anterior (ver Óptimo mediante Relajación), que aparecen sombreadas al lado de las hipótesis.

Si tuviéramos propiedades (o algoritmos) para encontrar el óptimo global del problema $(P_{\lambda,\mu})$, el TMFRL da ecuaciones adicionales que nos asegurarían que ese óptimo también vale para (P).

Suficiencia para un Óptimo (Karush-Kuhn-Tucker)

Teorema (Las condiciones KKT son suficientes)

En un problema (P) donde las funciones f y g_i son convexas, las h_j afines, y además todas son diferenciables, es suficiente encontrar $\bar{x}$, $\bar{\lambda}=\{\lambda_i\}$ y $\bar{\mu}=\{\mu_j\}$ que cumplan las condiciones KKT siguientes:

1. $\nabla f_{\bar{\lambda},\bar{\mu}}(\bar{x}) = \nabla f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\bar{x}) = \vec{0}$
2. $g_i(\bar{x}) \leq 0\,(1 \leq i \leq p)$ *y* $h_j(\bar{x}) = 0$
3. $\lambda_i \geq 0\,(1 \leq i \leq p)$
4. $\lambda_i \cdot g_i(\bar{x}) = 0\,(1 \leq i \leq p)$ *[condición de complementaridad]*

para tener un óptimo global $\bar{x}$ del problema.

PRUEBA: $f_{\lambda,\mu}(x)$ siempre es diferenciable (combinación lineal de diferenciables). Cuando $q = 0$ (i.e. $\{h_j\} = \emptyset$), sabemos que $f(x) + \sum_{i=1}^{p} \lambda_i g_i(x)$ es convexa porque $\lambda_i \geq 0$ (tercera condición).

Suficiencia para un Óptimo (Karush-Kuhn-Tucker)

Teorema (Las condiciones KKT son suficientes)

En un problema (P) donde las funciones f y g_i son convexas, las h_j afines, y además todas son diferenciables, es suficiente encontrar $\bar{x}$, $\bar{\lambda}=\{\lambda_i\}$ y $\bar{\mu}=\{\mu_j\}$ que cumplan las condiciones KKT siguientes:

1. $\nabla f_{\bar{\lambda},\bar{\mu}}(\bar{x}) = \nabla f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\bar{x}) = \vec{0}$
2. $g_i(\bar{x}) \leq 0 \, (1 \leq i \leq p)$ *y* $h_j(\bar{x}) = 0$
3. $\lambda_i \geq 0 \, (1 \leq i \leq p)$
4. $\lambda_i \cdot g_i(\bar{x}) = 0 \, (1 \leq i \leq p)$ *[condición de complementaridad]*

para tener un óptimo global $\bar{x}$ del problema.

PRUEBA: Luego, $\nabla f_{\bar{\lambda}}(\bar{x}) = \vec{0}$ (condición 1) es suficiente para un óptimo global del problema (P_λ), y el resto de las condiciones son las del TMFRL.

Suficiencia para un Óptimo (Karush-Kuhn-Tucker)

Teorema (Las condiciones KKT son suficientes)

En un problema (P) donde las funciones f y g_i son convexas, las h_j afines, y además todas son diferenciables, es suficiente encontrar $\bar{x}$, $\bar{\lambda}=\{\lambda_i\}$ y $\bar{\mu}=\{\mu_j\}$ que cumplan las condiciones KKT siguientes:

1. $\nabla f_{\bar{\lambda},\bar{\mu}}(\bar{x}) = \nabla f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\bar{x}) = \vec{0}$
2. $g_i(\bar{x}) \leq 0 \, (1 \leq i \leq p)$ y $h_j(\bar{x}) = 0$
3. $\lambda_i \geq 0 \, (1 \leq i \leq p)$
4. $\lambda_i \cdot g_i(\bar{x}) = 0 \, (1 \leq i \leq p)$ *[condición de complementaridad]*

para tener un óptimo global $\bar{x}$ del problema.

PRUEBA: Las funciones afines son las que preservan lineas y paralelismos (e.g. las funciones lineales son afines). Mantienen la convexidad, incluso al ser multiplicadas por negativos.

Suficiencia para un Óptimo (Karush-Kuhn-Tucker)

Teorema (Las condiciones KKT son suficientes)

En un problema (P) donde las funciones f y g_i son convexas, las h_j afines, y además todas son diferenciables, es suficiente encontrar $\bar{x}$, $\bar{\lambda}=\{\lambda_i\}$ y $\bar{\mu}=\{\mu_j\}$ que cumplan las condiciones KKT siguientes:

1. $\nabla f_{\bar{\lambda},\bar{\mu}}(\bar{x}) = \nabla f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\bar{x}) = \vec{0}$
2. $g_i(\bar{x}) \leq 0\,(1 \leq i \leq p)$ *y* $h_j(\bar{x}) = 0$
3. $\lambda_i \geq 0\,(1 \leq i \leq p)$
4. $\lambda_i \cdot g_i(\bar{x}) = 0\,(1 \leq i \leq p)$ *[condición de complementaridad]*

para tener un óptimo global $\bar{x}$ del problema.

PRUEBA: Si $q > 0$, pueden haber μ_j's< 0 y aun así sostenerse la convexidad de $f_{\lambda,\mu}(x)$ por la afinidad en las h_j, lo que completa la suficiencia de las condiciones.

Ejemplo KKT

Sea $(P)\begin{cases} \max x + y \\ x^2 + y^2 \leq 1 \end{cases}$, equivalente a $(P')\begin{cases} \min -(x+y) \\ x^2 + y^2 - 1 \leq 0 \end{cases}$.

Luego, $f_\lambda(x,y) = -x - y + \lambda(x^2 + y^2 - 1)$ y resolvemos el sistema:

i) $-1 + 2\lambda x = 0$ y $-1 + 2\lambda y = 0$,

ii) $x^2 + y^2 - 1 \leq 0$ (equivalente a $x^2 + y^2 \leq 1$),

iii) $\lambda \geq 0$

iv) $\lambda(x^2 + y^2 - 1) = 0$.

Si $\lambda = 0$ no puede cumplirse i). Entonces $\lambda > 0$ (por iii), $x^2 + y^2 = 1$ (por iv y cumple ii) y $x = y = \frac{1}{2\lambda}$ (por i).
Substituyendo x y y llegamos a $(\frac{1}{2\lambda})^2 + (\frac{1}{2\lambda})^2 = 1 = \frac{1}{2\lambda^2}$, de donde $\lambda = \frac{1}{\sqrt{2}}$, $\bar{x} = \bar{y} = \frac{1}{\sqrt{2}}$ es el óptimo y $f(\bar{x},\bar{y})$ vale $-\sqrt{2}$ en (P').
El óptimo en (P) se da en el mismo punto, pero vale $\sqrt{2}$.

Ejemplo KKT

Sea $(P)\begin{cases} \max x+y \\ x^2+y^2 \leq 1 \end{cases}$, equivalente a $(P')\begin{cases} \min -(x+y) \\ x^2+y^2-1 \leq 0 \end{cases}$.

Luego, $f_\lambda(x,y) = -x-y+\lambda(x^2+y^2-1)$ y resolvemos el sistema:

- i) $-1+2\lambda x = 0$ y $-1+2\lambda y = 0$,
- ii) $x^2+y^2-1 \leq 0$ (equivalente a $x^2+y^2 \leq 1$),
- iii) $\lambda \geq 0$
- iv) $\lambda(x^2+y^2-1) = 0$.

¿Qué hubiera pasado si no hubiésemos encontrado solución al sistema? ¿Podemos decir que el problema no tiene óptimo?

En este caso tiene óptimo por Weierstrass. En general ¿estamos seguros que KKT se cumple? ¿podemos asegurar que no hay óptimo si sus ecuaciones no tienen solución?

Suficiencia para un Óptimo (Karush-Kuhn-Tucker)

Las condiciones de KKT son en general para clasificar óptimos locales. En el caso convexo, se clasifican ambos tipos simultáneamente y no hacemos la diferencia.

Pero las condiciones KKT son suficientes para clasificar óptimos locales incluso bajo hipótesis más generales, como que las funciones sean Invexas: $f : \mathbb{R}^n \to \mathbb{R}$ diferenciable es invexa sii para todo $x, y \in \mathbb{R}^n$ existe $\eta(x, y)$ tal que $f(y) - f(x) \geq \nabla^T f(x) \cdot \eta(x, y)$. Cuando $\eta(x, y) = y - x$ es la convexidad.

Suficiencia para un Óptimo (Karush-Kuhn-Tucker)

Las condiciones de KKT son en general para clasificar óptimos locales. En el caso convexo, se clasifican ambos tipos simultáneamente y no hacemos la diferencia.

Pero las condiciones KKT son suficientes para clasificar óptimos locales incluso bajo hipótesis más generales, como que las funciones sean Invexas. **No las vemos en este curso.**

En la prueba del TM previo se exigió que las funciones f y g_i sean convexas, y en caso de existir las h_j, que ellas sean afines. En esas hipótesis se cumple la Invexidad

Otra forma general de garantizar la suficiencia de un óptimo local la constituyen las *Second Order Sufficient Conditions* (SOSC), que generalizan la idea de que la Hessiana sea semidefinida positiva. **Tampoco las vemos en el curso.**

Otro Caso (optimización con igualdades)

Cuando no hay desigualdades (i.e. $\{g_i\}=\emptyset$), el problema queda:

$$(P)\begin{cases} \min f(x) \\ h_j(x) = 0, \quad 1 \leq j \leq q \end{cases},$$

y las condiciones KKT son las condiciones de Lagrange:

$$\begin{cases} \nabla f_{\bar{\mu}}(\bar{x}) = \nabla f(\bar{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\bar{x}) = \vec{0} \\ h_j(\bar{x}) = 0, \ 1 \leq j \leq q \end{cases}$$

Pensemos en un caso simple en $\mathbb{R}^2$, con una única restricción:

$$(P)\begin{cases} f(x,y) = x^2 - 2x + 4y^2 \\ xy = 4 \ \text{ (i.e. } g(x,y) = xy - 4 = 0) \end{cases}.$$

Otro Caso (optimización con igualdades)

Pensemos en un caso simple en $\mathbb{R}^2$, con una única restricción:

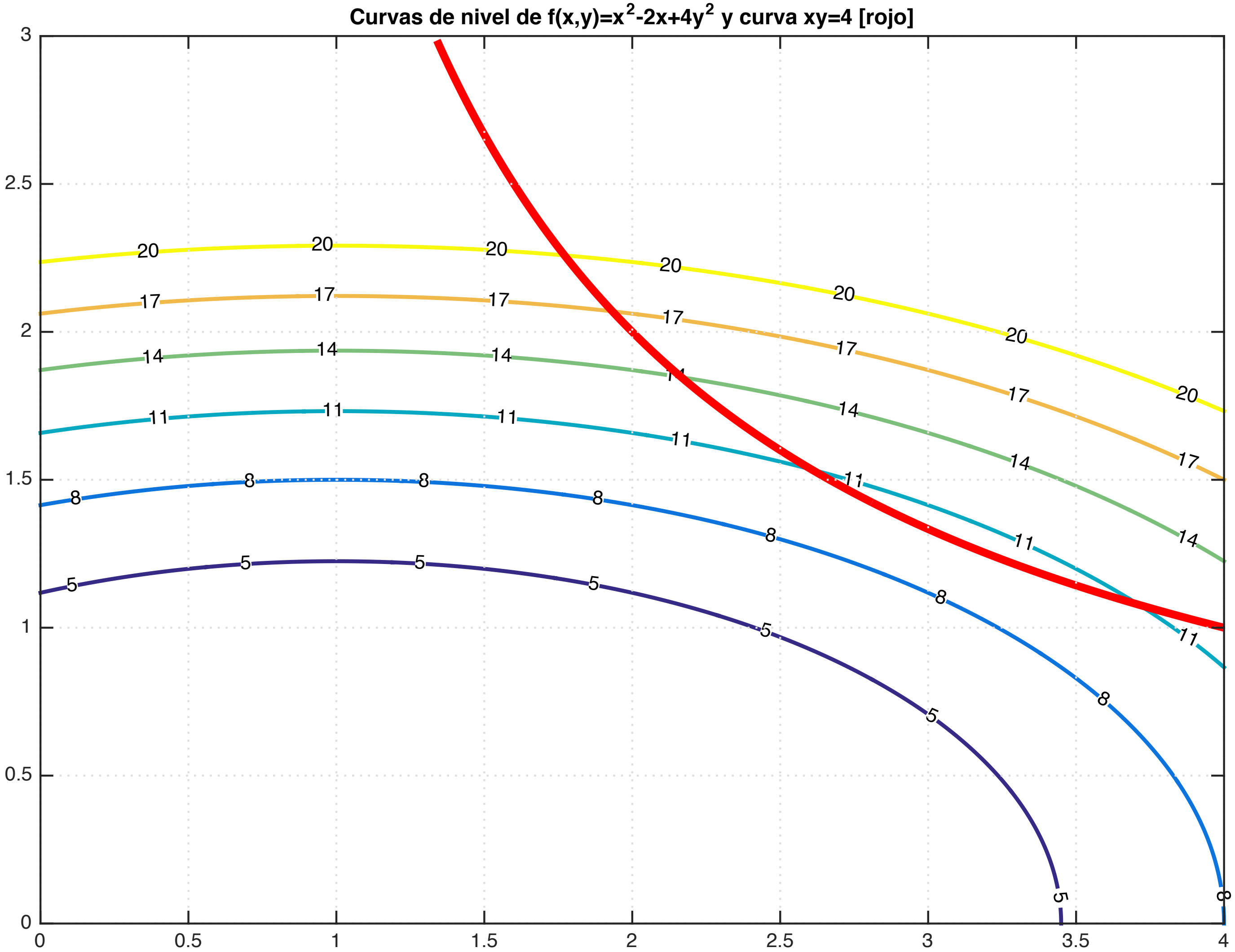
$$(P) \begin{cases} f(x,y) = x^2 - 2x + 4y^2 \\ xy = 4 \ \text{(i.e. } g(x,y) = xy - 4 = 0) \end{cases}.$$

Las ecuaciones quedan: $\nabla f_\mu(x,y) = \begin{bmatrix} 2x-2 \\ 8y \end{bmatrix} + \mu \begin{bmatrix} y \\ x \end{bmatrix} = \vec{0}$, con $xy = 4$. Luego $y = \frac{4}{x}$, $\frac{32}{x} + \mu x = 0 \Rightarrow \mu = -\frac{32}{x^2}$. Finalmente, $2x - 2 - \frac{32}{x^2} \cdot \frac{4}{x} = 0$, así que $x^3(x-1) - 64 = 0$.

La raíz está entre 3 y 4. Podemos usar Newton, y conseguimos la secuencia $x_0 = 3$, $x_1 = \frac{253}{81}$, $x_2 = \frac{2717}{872}$, $x_3 = \frac{6431}{2064}$. La diferencia entre los últimos es $3.1e^{-5}$, así que $x \approx 3.1158$, $y \approx 1.2838$, $\mu \approx -3.2962$ y el óptimo vale $f(\bar{x},\bar{y}) \approx 10.069$.

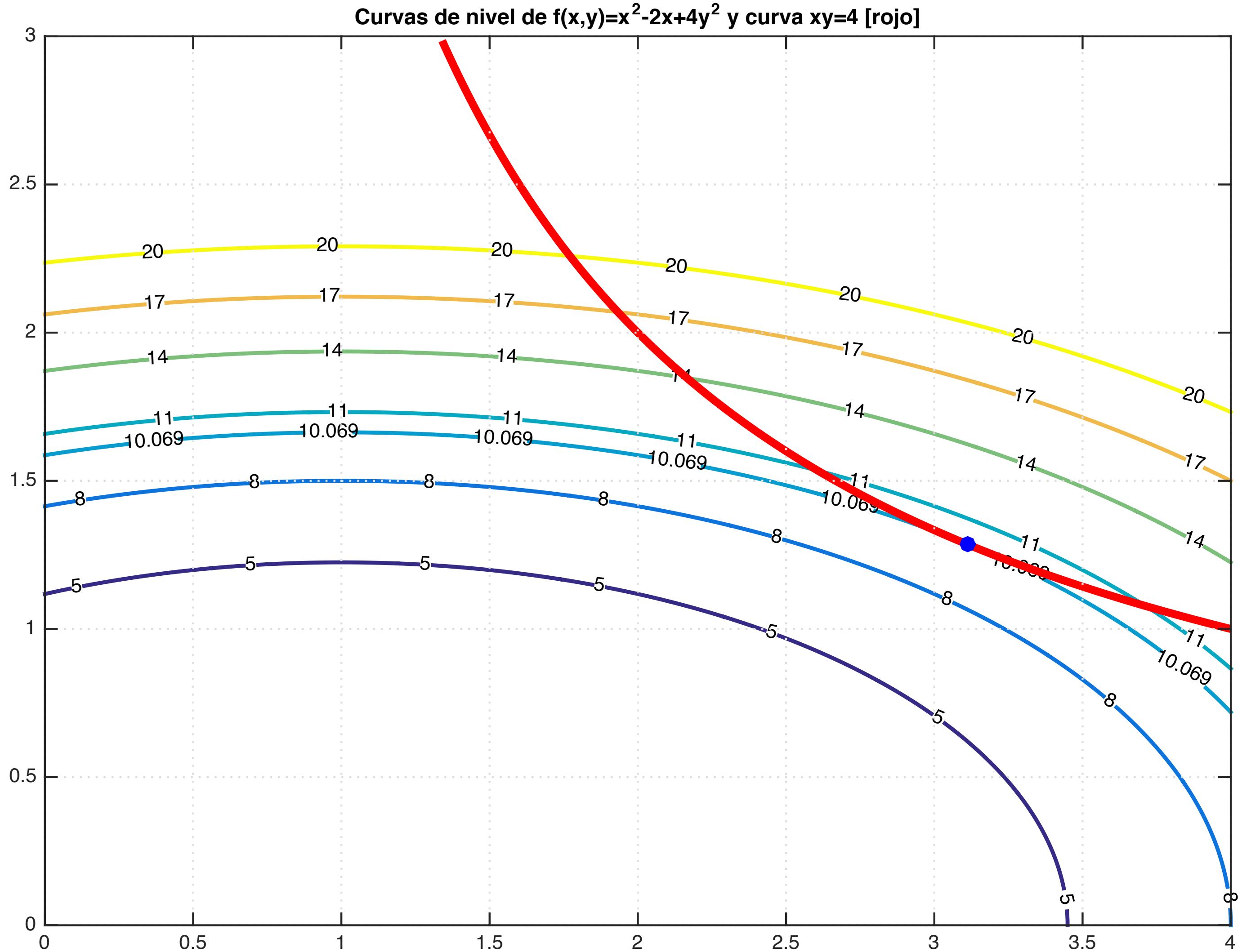
Otro Caso (optimización con igualdades)

¿Cómo podríamos reconocer la solución intuitivamente?

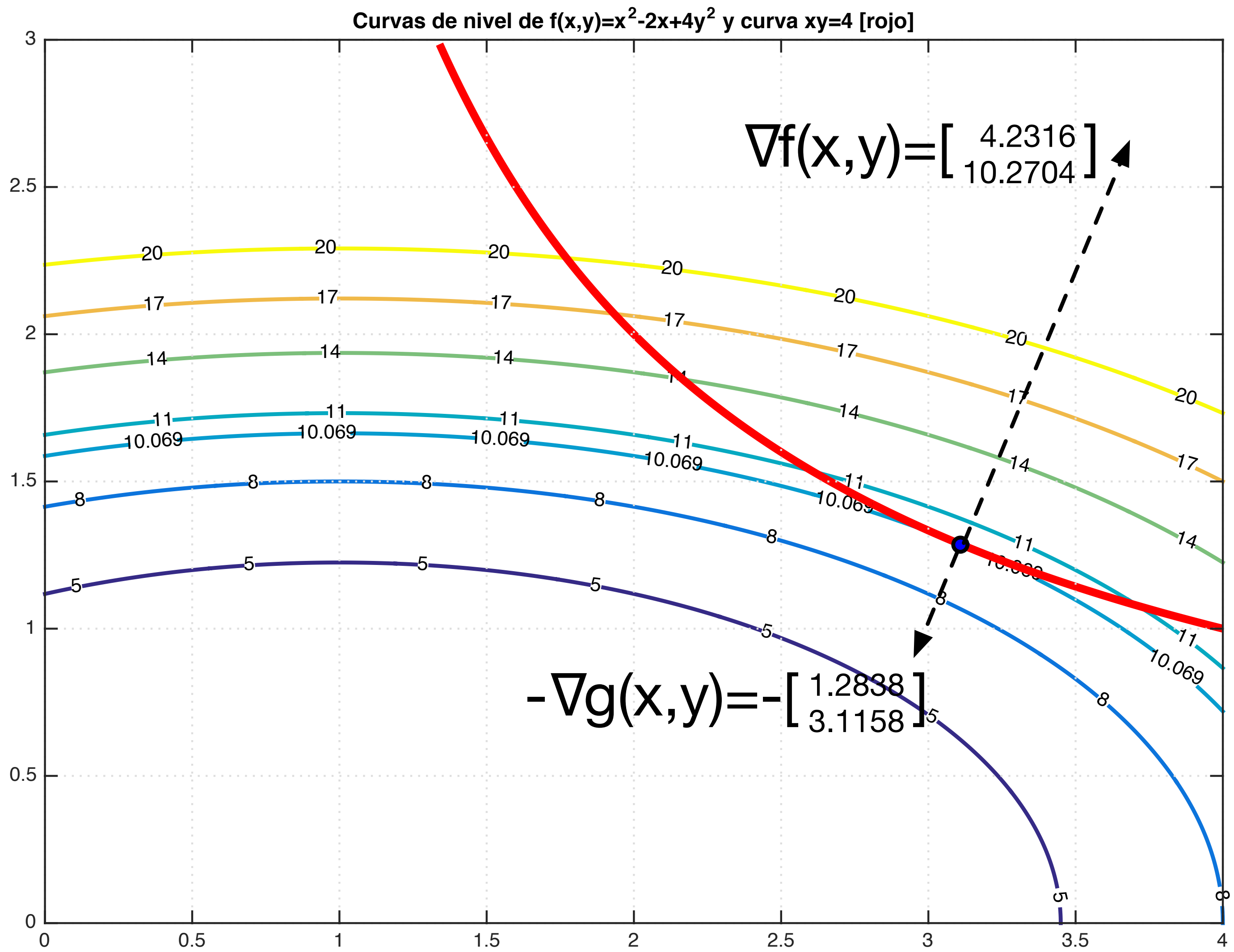


Otro Caso (optimización con igualdades)

¿Cómo podríamos reconocer la solución intuitivamente?



Otro Caso (optimización con igualdades)



Buscando que las normales a $xy=4$ y $f(x,y)$ sean colineales.

Necesidad de las condiciones KKT

Que las condiciones KKT sean necesarias se sustenta en la geometría de la región factible, más que en argumentos del cálculo.

Pensemos en el caso con desigualdades (el otro es Lagrange).

Si el óptimo se alcanza en el interior de la región factible, el problema es *sin restricciones* luego $\bar{\lambda} = \vec{0}$ y $\nabla f(\bar{x}) = \vec{0}$, entonces estamos como en la primera parte del curso.

Estudiamos el caso en que $\bar{x}$ está en el frontera.

Definición (Cono Normal en un $\bar{x}$ de la frontera)

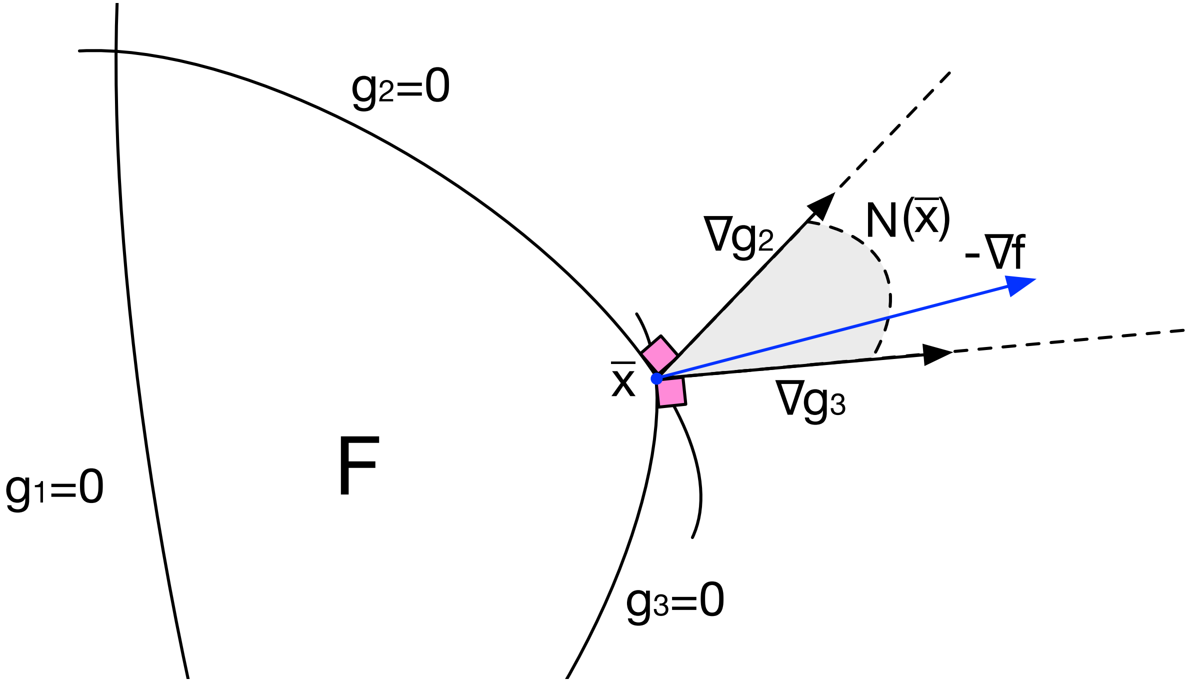
Dado (P) y $\bar{x} \in F$ (su reg. fact.) tal que $I_a \subseteq \{1, \ldots, p\}$ es el conjunto donde $g_i(\bar{x}) = 0$ sii $i \in I_a$ (i.e. restricciones activas). Llamamos $\mathcal{N}(\bar{x}) = \{v \in \mathbb{R}^n : v = \sum_{i \in I_a} \alpha_i \nabla g_i(\bar{x}),\ con\ \alpha_i \geq 0\}$

Observar que $g_i(\bar{x}) < 0$ si $i \notin I_a$ porque $\bar{x}$ es factible en (P).

Necesidad de las condiciones KKT

Definición (Cono Normal en un $\bar{x}$ de la frontera)

Dado (P) y $\bar{x} \in F$ (su reg. fact.) tal que $I_a \subseteq \{1, \ldots, p\}$ es el conjunto donde $g_i(\bar{x}) = 0$ sii $i \in I_a$ (i.e. restricciones activas).
Llamamos $\mathcal{N}(\bar{x}) = \{v \in \mathbb{R}^n : v = \sum_{i \in I_a} \alpha_i \nabla g_i(\bar{x}),\ \text{con } \alpha_i \geq 0\}$



Necesidad de las condiciones KKT

Definición (Cono Normal en un $\bar{x}$ de la frontera)

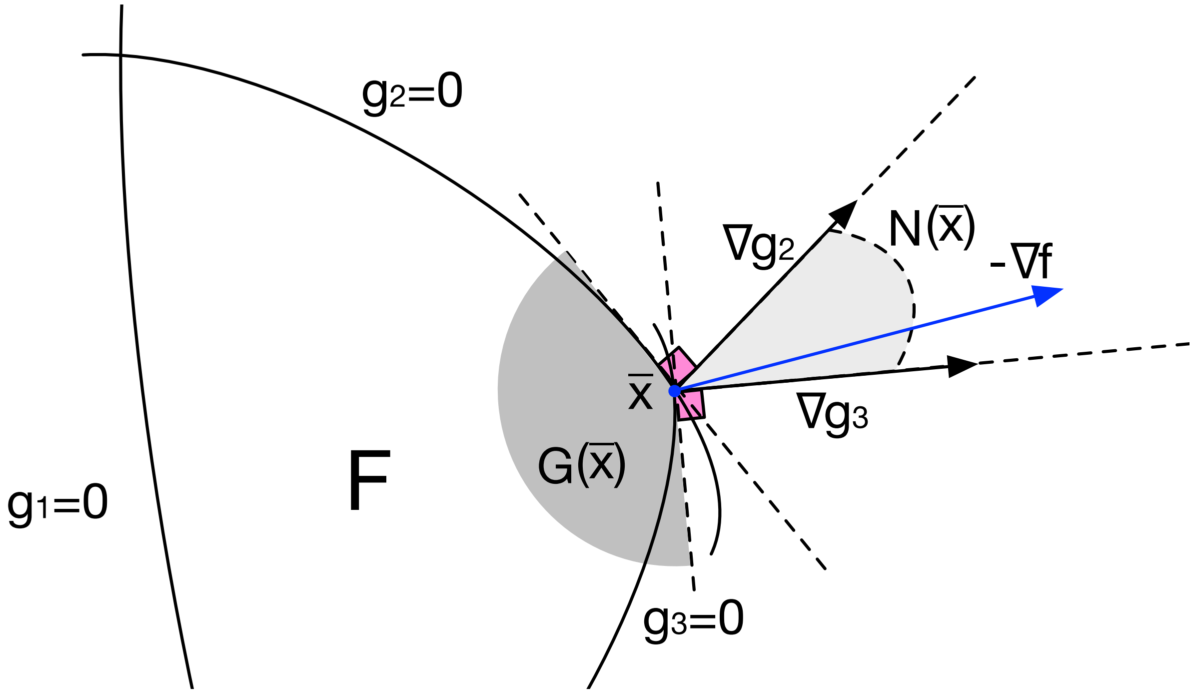
Dado (P) y $\bar{x} \in F$ (su reg. fact.) tal que $I_a \subseteq \{1, \ldots, p\}$ es el conjunto donde $g_i(\bar{x}) = 0$ sii $i \in I_a$ (i.e. restricciones activas). Llamamos $\mathcal{N}(\bar{x}) = \{v \in \mathbb{R}^n : v = \sum_{i \in I_a} \alpha_i \nabla g_i(\bar{x}),$ con $\alpha_i \geq 0\}$

La condición $\nabla f_{\bar{\lambda}}(\bar{x}) = \vec{0}$ con $\bar{\lambda} \geq 0$ equivale a que la dirección de máximo decrecimiento de f en $\bar{x}$ esté en $\mathcal{N}(\bar{x})$.

Se cumple que el cono $\mathcal{G}(\bar{x}) = \{v \in \mathbb{R}^n : \nabla^T g_i(\bar{x}) \cdot v \leq 0, \forall i \in I_a\}$ (el Cono Dual) es el ortogonal a $\mathcal{N}(\bar{x})$, esto es $\mathcal{G}(\bar{x}) = \mathcal{N}^{\perp}(\bar{x})$.

Necesidad de las condiciones KKT

Se cumple que el cono $\mathcal{G}(\bar{x}) = \{v \in \mathbb{R}^n : \nabla^T g_i(\bar{x}) \cdot v \leq 0, \forall i \in I_a\}$ (el Cono Dual) es el ortogonal a $\mathcal{N}(\bar{x})$, esto es $\mathcal{G}(\bar{x}) = \mathcal{N}^{\perp}(\bar{x})$.



Para que KKT sea necesario, basta que la tangente a cualquier trayectoria originada en $\bar{x}$ esté contenida en $\mathcal{G}(\bar{x})$: que las direcciones para decrecer sean dentro del cono normal, fuera de F.

Necesidad de las condiciones KKT

Definición (Cono Tangente en un punto $\bar{x}$ de la frontera)

Dado (P) de región factible F no-trivial y $\bar{x}$ en su borde, llamamos $\mathcal{T}(\bar{x}) = \{\vec{d} \in \mathbb{R}^n : \vec{d}$ tangente a una curva admisible desde $\bar{x}\}$ al Cono Tangente en $\bar{x}$. Decimos que el punto $\bar{x}$ es Regular en (P) cuando $\mathcal{T}(\bar{x}) = \mathcal{G}(\bar{x})$.

Lema (Farkas-Minkowski)

Sea $A \in \mathbb{R}^{m \times n}$ no singular cualquiera, con $n > m$ y $b \in \mathbb{R}^m$. Se cumple que $\exists x \geq 0$ $(x \in \mathbb{R}^n)$, tal que $Ax = b$ si y sólo si, $b^T u \geq 0$ para todo $u \in \mathbb{R}^m$ tal que $A^T u \geq 0$.

Lo probaremos como parte del práctico, con argumentos puramente algebraicos en el problema de Programación Lineal.

Necesidad de las condiciones KKT

Teorema (Las condiciones KKT son necesarias)

En un problema (P) de región factible no-trivial, donde las funciones f y g_i son diferenciables, si existe un óptimo local de (P) en $\bar{x}$, punto regular en la frontera de (P), las condiciones KKT deben cumplirse.

PRUEBA: Sabemos que $g_i(\bar{x}) \leq 0$ porque $\bar{x} \in F$. Hay que probar que existe $\bar{\lambda} = \{\lambda_i\}$ que cumple KKT. Si $i \notin I_a$ (no activa), tomo $\lambda_i = 0$ y nos concentramos en las activas.

Cualquier curva admisible puede representarse como $x : [0, 1) \to \mathbb{R}^n$ diferenciable donde: $x(t) \in F, t \in [0, 1)$ y $x(0) = \bar{x}$. Como $\bar{x}$ es óptimo local, debe verificarse que $h'(0) \geq 0$, donde $h(t) = f(x(t))$, esto es: la función f no puede decrecer sobre curvas admisibles.

Necesidad de las condiciones KKT

Teorema (Las condiciones KKT son necesarias)

En un problema (P) de región factible no-trivial, donde las funciones f y g_i son diferenciables, si existe un óptimo local de (P) en $\bar{x}$, punto regular en la frontera de (P), las condiciones KKT deben cumplirse.

PRUEBA: Luego, $h'(0) = \nabla^T f(x(0)) \cdot x'(0) = \nabla^T f(\bar{x}) \cdot \vec{d} \geq 0$, para todo $\vec{d} \in \mathcal{T}(\bar{x}) = \mathcal{G}(\bar{x})$ porque $\bar{x}$ es regular.
Luego $\nabla^T f(\bar{x}) \cdot \vec{d} \geq 0$ para todo $\vec{d}$ tal que $\nabla^T g_i(\bar{x}) \cdot \vec{d} \leq 0, i \in I_a$.
Tomo $b = \nabla f(\bar{x})$ y $A = [\cdots | -\nabla g_i(\bar{x}) | \cdots]$ y se cumple que:
$b^T \cdot \vec{d} \geq 0$ para todo $\vec{d}$ tal que $A^T \cdot \vec{d} \geq 0$.
Por Farkas-Minkowski, existe $\bar{\alpha} \geq 0$ $(\bar{\alpha} = \{\alpha_i\})$ tal que $A \cdot \bar{\alpha} = b$, que equivale a: $\nabla f(\bar{x}) = -\sum_{i \in I_a} \alpha_i \nabla g_i(\bar{x})$.

Necesidad de las condiciones KKT

Teorema (Las condiciones KKT son necesarias)

En un problema (P) de región factible no-trivial, donde las funciones f y g_i son diferenciables, si existe un óptimo local de (P) en $\bar{x}$, punto regular en la frontera de (P), las condiciones KKT deben cumplirse.

PRUEBA: Completando $\bar{\lambda}$ (era $\lambda_i = 0$ si $i \notin I_a$) con $\lambda_i = \alpha_i$ cuando $i \in I_a$, concluimos que $\nabla f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\bar{x}) = \vec{0}$ con $\lambda_i \geq 0, 1 \leq i \leq p$.

Eso prueba las ecuaciones 1 a 3 de KKT. La complementaridad (ecuación 4) se cumple por construcción, ya que $\lambda_i = 0$ si $i \notin I_a$ (i.e. si $g_i(\bar{x}) < 0$), y si $\lambda_i > 0$ es porque $i \in I_a$ y $g_i(\bar{x}) = 0$.

Necesidad de las condiciones KKT

Teorema (Las condiciones KKT son necesarias)

En un problema (P) de región factible no-trivial, donde las funciones f y g_i son diferenciables, si existe un óptimo local de (P) en $\bar{x}$, punto regular en la frontera de (P), las condiciones KKT deben cumplirse.

Así como la Invexidad y las Second Order Sufficient Conditions garantizaban suficiencia de las condiciones KKT, hay condiciones de regularidad (llamadas *constraint qualifications*) para las cuales las condiciones KKT son necesarias.

Tampoco las veremos en este curso porque las regiones factibles convexas no-triviales las cumplen.

Necesidad y suficiencia de las condiciones KKT (convexidad)

Teorema (Las condiciones KKT son necesarias en el caso convexo)

En un problema (P) donde las f y g_i son convexas, las h_j afines, F es no-trivial, y las funciones son diferenciables, las condiciones KKT son necesarias y suficientes para la existencia de óptimos.

La suficiencia ya la vimos. La necesidad se debe a que los dominios convexos no triviales cumplen las condiciones de regularidad, y el dominio F de (P) es la intersección de convexos, ya que las restricciones $g_i(x) \leq 0$ definen conjuntos convexos (g_i's son convexas), las $h_j(x) = 0$ también (son afines) y la región factible es entonces intersección de conexos.

Que el conjunto sea no-trivial equivale a que exista $\hat{x}$ tal que $g_i(\hat{x}) < 0$ con $i = 1, \ldots, p$, **y debe cumplirse para garantizar que las hipótesis del teorema se cumplen.**