programa lineal: $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m-n \times n}$, $b \in \mathbb{R}^{m-n}$, $m > n$

$$\left\{\begin{array}{ll} \min\limits_{x} & c^T x \\ \text{s.a.} & Ax \leq b \\ & x \geq 0 \end{array}\right.$$

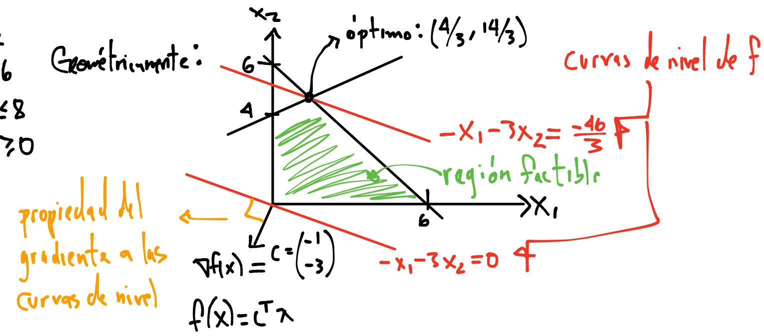
ejemplo: $\min -x_1 - 3x_2$

s.a. $x_1 + x_2 \leq 6$

$-x_1 + 2x_2 \leq 8$

$x_1, x_2 \geq 0$

Geométricamente:



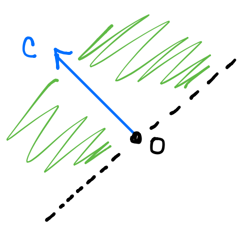
Teorema de Farkas:

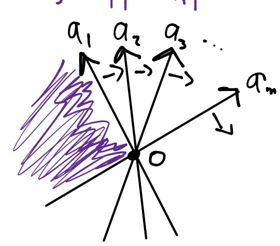
Uno y sólo uno de los 2 sistemas siguientes tiene una solución: 1) $c^T p > 0$, $Ap \leq 0$ 2) $A^T \lambda = c$, $\lambda \geq 0$

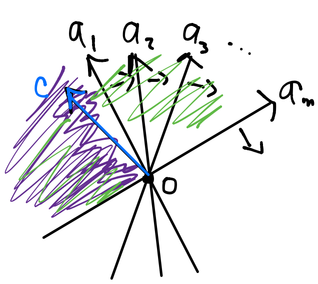
Interpretación geométrica: $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m-n \times n}$ entonces:

Sistema 1 tiene solución: el semiespacio abierto $\{p : c^T p > 0\}$ intersección con el cono $\{p : Ap \leq 0\}$ no es vacía:

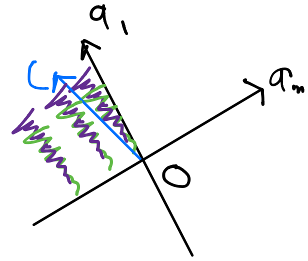
$\{p : c^T p > 0\}$ $\{p : Ap \leq 0\} = \{p : a_1^T p \leq 0, a_2^T p \leq 0 \ldots a_m^T p \leq 0\}$



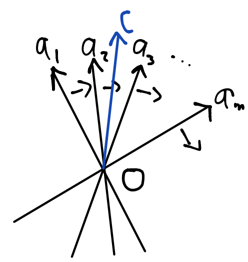




intersección:



Sistema 2 tiene solución si y sólo si c pertenece al cono generado por los renglones de A:



Obs: Sólo un sistema tiene solución (si uno tiene solución el otro no tiene y viceversa)

Condiciones de KKT para el problema $\min_x c^T x$ s.a. $Ax \leq b$, $-x \leq 0$ $\Big\}$ $\min c^T x$ s.a $\begin{bmatrix} A \\ -I \end{bmatrix} x \leq \begin{bmatrix} b \\ 0 \end{bmatrix}$ $\Big\}$ $\min c^T x$ s.a. $\tilde{A}x \leq \tilde{b} \in \mathbb{R}^m$

$$\mathcal{L}(x,\tilde{\lambda}) = c^T x + \tilde{\lambda}^T(\tilde{A}x - \tilde{b}) = c^T x + \underbrace{\lambda_1^T}_{\lambda_1 \in \mathbb{R}^m}(Ax-b) + \underbrace{\lambda_2^T}_{\lambda_2 \in \mathbb{R}^n}(-x)$$

Condiciones de KKT: x mínimo ent. $\exists \lambda_1, \lambda_2$ que satisfacen:

$$\nabla_x \mathcal{L}(x,\lambda_1,\lambda_2) = 0 = c + A^T\lambda_1 - \lambda_2 \rightsquigarrow \nabla f(x) = -\sum_{i=1}^{m} \lambda_1(i) a_i - \sum_{i=1}^{n} \lambda_2(i) e_i$$

$$\left.\begin{aligned} \lambda_1^T(Ax-b) &= 0 \\ \lambda_2^T(-x) &= 0 \end{aligned}\right\} \text{holgura complementaria}$$

$$Ax - b \leq 0$$
$$-x \leq 0$$
$$\tilde{\lambda} \geq 0$$

gradiente de f es combinación lineal de $-a_i$, $-e_i$. De hecho por holgura complementaria es combinación lineal únicamente de las restricciones activas (pues las inactivas, $f_i(x) < 0$ tienen $\lambda_i = 0$ para alguna $i = 1, \ldots m$)

Problema dual: $\max -b^T\lambda_1$

$$\text{s.a.}\quad c + A^T\lambda_1 - \lambda_2 = 0$$
$$\lambda_1 \geq 0$$
$$\lambda_2 \geq 0$$

¿Relación entre condiciones KKT y el teorema de Farkas?

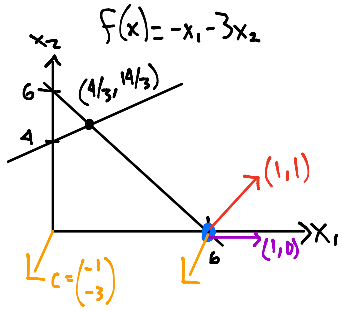
Si x no es mínimo entonces $\exists$ dirección de descenso Δx. Supóngase que $A\Delta x \geq 0$ entonces:

1) $\nabla f(x)^T \Delta x = c^T \Delta x < 0$ (dirección de descenso) 2) $A\Delta x \geq 0$ (suposición)

$\therefore$ $\begin{matrix} c^Tp > 0 \\ Ap \leq 0 \end{matrix}$ tiene solución (tomando $p = -\Delta x$) y por Teorema de Farkas $\begin{matrix} A^T\lambda = c \\ \lambda \geq 0 \end{matrix}$

no tiene solución (gradiente de f objetivo no es combinación lineal de cada gradiente de las restricciones activas)

En el ejemplo:
$$\begin{aligned} \min\ & -x_1 - 3x_2 \\ \text{s.a.}\ & x_1 + x_2 \leq 6 \\ & -x_1 + 2x_2 \leq 8 \\ & x_1, x_2 \geq 0 \end{aligned}$$
se tiene:



(6,0) no es mínimo de f $\therefore$ c no es combinación lineal del gradiente de las restricciones activas:

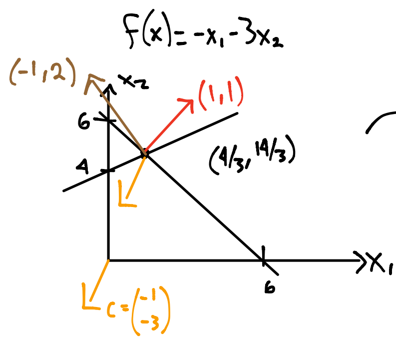
$$c = \begin{pmatrix} -1 \\ -3 \end{pmatrix} \neq \lambda_1^{(1)}\left(-\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) + \lambda_2^{(1)}\left(-\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right),\quad \begin{matrix} \lambda_1^{(1)} \geq 0 \\ \lambda_2^{(1)} \geq 0 \end{matrix}$$ pero sí se cumple

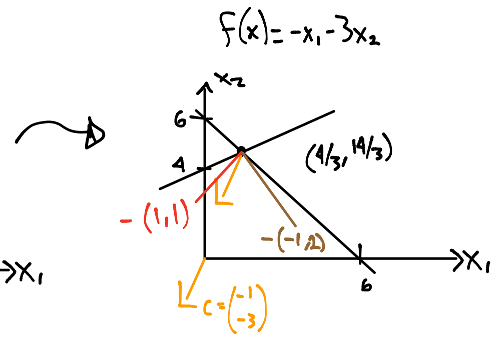
$c^Tp > 0$, $Ap \leq 0$ tiene solución $\therefore$ $\exists$ dirección de descenso

Mismo análisis para los puntos $(0,0)$, $(0,4)$. Para el mínimo $(4/3, 14/3)$ las restricciones activas son $\begin{matrix} x_1 + x_2 \leq 6 \\ -x_1 + 2x_2 \leq 8 \end{matrix}$ y c es combinación lineal del gradiente de las restricciones activas ($\begin{matrix} A^T\lambda = c \\ \lambda \geq 0 \end{matrix}$ sí tiene solución)

$c = \begin{pmatrix}-1\\-3\end{pmatrix} = \lambda_1\left(-\begin{pmatrix}1\\1\end{pmatrix}\right) + \lambda_2\left(-\begin{pmatrix}-1\\2\end{pmatrix}\right), \lambda_1 = \frac{5}{3}, \lambda_2 = \frac{2}{3}$. El problema $c^Tp > 0,\ Ap \le 0$ no tiene solución

$\therefore \nexists$ dirección de descenso :





Para problemas $\min f_0(x)$ con f_0 y h_i convexas y diferenciables $\forall i = 1, \dots p$

sujeto a: $h_i(x) = 0$

$i = 1, \dots, p$

si x no es mínimo entonces $\exists\ \Delta x$ dirección de descenso y suponemos Δx satisface

$\nabla h_i(x)^T \Delta x = 0\ \ \forall i = 1, \dots, p$ entonces reescribiendo $\nabla h_i(x)^T \Delta x = 0\ \ \forall i = 1, \dots, p$

como $\begin{matrix}\nabla h_i(x)^T \Delta x \le 0 \\ \nabla h_i(x)^T \Delta x \ge 0\end{matrix}$ $\forall i = 1, \dots, p$ se tiene: 1) $\underbrace{\nabla f_0(x)^T \Delta x < 0}_{\text{dirección de descenso}}$ 2) $\underbrace{\begin{matrix}\nabla h_i(x)^T \Delta x \ge 0 \\ -\nabla h_i(x)^T \Delta x \ge 0 \\ \forall i = 1, \dots, p\end{matrix}}_{\text{suposición}}$

$\therefore \begin{matrix}\nabla f_0(x)^T p > 0 \\ Ap \le 0\end{matrix}$ tiene solución $\left(\text{tomando } p = -\Delta x \text{ y } A = \begin{bmatrix}\nabla h_1(x) \\ \vdots \\ \nabla h_p(x)\end{bmatrix}\right)$ y por Teorema de Farkas $\begin{matrix}A^T \lambda = \nabla f_0(x) \\ \lambda \ge 0\end{matrix}$

no tiene solución (gradiente de f_0 no es combinación lineal de cada gradiente de las restricciones activas)

Asimismo, x mínimo local de f $\quad \nabla f_0(x)^T p > 0$ no tiene solución y por Teorema de Farkas
$$Ap \leq 0$$

$$A^T\lambda = \nabla f_0(x)$$
$$\lambda \geq 0$$
tiene solución (gradiente de f_0 es combinación lineal de cada gradiente de las restricciones activas)

(Mismo análisis se realiza para
$$\min f_0(x)$$
$$\text{sujeto a: } f_i(x) \leq 0 \quad \forall i = 1, \dots, m$$
) ... Entonces:

$$\min f_0(x)$$
$$\text{sujeto a: } f_i(x) \leq 0 \quad \forall i=1,\dots,m$$
$$h_i(x) = 0 \quad \forall i=1,\dots,p$$
se define la Lagrangiana como una función

$L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \longrightarrow \mathbb{R}$ asociada al problema anterior como:

$$\mathcal{L}(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \quad \text{con } \operatorname{dom} L = D \times \mathbb{R}^m \times \mathbb{R}^p.$$

Notas:
- λ_i se llama multiplicador de Lagrange asociado con la i-ésima restricción de desigualdad $f_i(x) \leq 0$.
- ν_i se llama multiplicador de Lagrange asociado con la i-ésima restricción de igualdad $h_i(x) = 0$.
- Los vectores λ, ν son llamados variables duales o vectores de multiplicadores de Lagrange asociados con el problema anterior.

Condiciones de Karush–Kuhn-Tucker (KKT) de optimalidad.

Supóngase que $f_0, f_1, \dots, f_m, h_1, h_2, \dots, h_p$ son diferenciables con dominios abiertos

Condiciones KKT para problemas no convexos.

Sean $x^*, (\lambda^*, \nu^*)$ un par primal-dual de puntos óptimos con duality gap cero.

Como x^* minimiza $L(x, \lambda^*, \nu^*)$ sobre x entonces:

$$\nabla_x L(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$

y por factibilidad y holgura complementaria:

$$f_i(x^*) \leq 0 \quad \forall\, i=1,\dots,m$$

$$\lambda_i^* \geq 0 \quad \forall\, i=1,\dots,m$$

$$h_i(x^*) = 0 \quad \forall\, i=1,\dots,p$$

$$\lambda_i^* f_i(x^*) = 0 \quad \forall\, i=1,\dots,m$$

Comentario: las 5 condiciones anteriores son llamadas las condiciones de optimalidad de Karush-Kuhn-Tucker que se satisfacen para cualquier problema de optimización con función objetivo y funciones restricción diferenciables evaluadas en un par primal dual de puntos óptimos en el que se cumple la dualidad fuerte. Obsérvese que son condiciones necesarias.

Condiciones KKT para problemas convexos.

Si el problema primal es convexo, esto es, f_i son convexas $\forall i=0,1,\dots,m$ h_i son afín $\forall i=1,\dots,p$ entonces las condiciones KKT de optimalidad son suficientes para puntos primal-dual óptimos:

Sean $\tilde{x},\tilde{\lambda},\tilde{\nu}$ puntos que satisfacen las condiciones KKT de optimalidad:

$$\nabla_x L(\tilde{x},\tilde{\lambda},\tilde{\nu}) = f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0$$

$$f_i(\tilde{x}) \leq 0 \quad \forall i=1,2,\dots,m \qquad h_i(\tilde{x}) = 0 \quad \forall i=1,2,\dots,p$$

$$\tilde{\lambda}_i \geq 0 \quad \forall i=1,2,\dots,m \qquad \tilde{\lambda}_i f_i(\tilde{x}) = 0 \quad \forall i=1,2,\dots,m$$

Entonces como se tiene un problema convexo, $\tilde{\lambda}_i \geq 0$ $\forall i=1,2,\dots,p$ y $\tilde{x},\tilde{\lambda},\tilde{\nu}$ satisfacen $\nabla_x L(\tilde{x},\tilde{\lambda},\tilde{\nu}) = 0$ resulta:

- $L(x,\tilde{\lambda},\tilde{\nu})$ es convexa
- $\tilde{x}$ es mínimo de $L(x,\tilde{\lambda},\tilde{\nu})$ y es primal factible.
- $g(\tilde{\lambda},\tilde{\nu}) = L(\tilde{x},\tilde{\lambda},\tilde{\nu}) = f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) = f_0(\tilde{x})$

por factibilidad primal y holgura complementaria

por lo tanto, $\tilde{x},(\tilde{\lambda},\tilde{\nu})$ tienen duality gap igual a cero y son puntos primal dual óptimos.

En resumen:

1) Para cualquier problema de optimización convexa con funciones objetivo y de restricciones diferenciables en el que una pareja de puntos primal-dual satisfacen las condiciones KKT de optimalidad, son óptimos para el problema primal y dual y tienen duality gap igual a cero.

2) Si un problema de optimización convexa con funciones objetivo y de restricción diferenciables satisface la condición de Slater entonces las condiciones de KKT proveen condiciones <u>necesarias y suficientes</u> para optimalidad de puntos primal-dual:

la condición de Slater implica que la duality gap es cero y el óptimo dual se alcanza, entonces x es punto óptimo primal si y sólo si existe (λ, ν) tal que junto con x satisfacen las condiciones KKT de optimalidad.

<u>Comentario</u>: en algunos casos especiales es posible resolver las condiciones KKT de manera analítica y en general muchos algoritmos de optimización resultan o pueden interpretarse como métodos que resuelven las condiciones de KKT.

La condición $\lambda_i^* f_i(x^*)=0$ se llama <u>holgura complementaria</u> y se cumple para cualquier punto óptimo primal x^* y punto óptimo dual (λ^*, ν^*) bajo dualidad fuerte.

La condición de holgura complementaria se escribe como: $\lambda_i^* > 0 \Rightarrow f_i(x^*)=0$

o equivalentemente: $f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$ esto es, el i-ésimo multiplicador de Lagrange es positivo entonces la i-ésima función restricción de desigualdad es activa.

Nota 4.5.3: $\min f(x)$ s.a $Ax=b$ Suposiciones:

1) $f:\mathbb{R}^n \to \mathbb{R}$ convexa y C^2 en su dom$f$
2) $A\in\mathbb{R}^{p\times n}$ rank$(A)=p<n$
3) $\exists$ x^* óptimo, p^* valor óptimo

Condiciones KKT: $x^* \in$ dom$f$ es óptimo si y sólo si:

$$\exists\, \nu \in \mathbb{R}^p \text{ tal que: } \underbrace{Ax^*=b}_{\text{factibilidad primal}},\ \underbrace{\nabla f(x^*)+A^T\nu^*=0}_{\text{factibilidad dual}}$$

2 condiciones y una idea basada en aproximación de 2° orden para extender el método de Newton a problemas de la forma $\min f(x)$ s.a $Ax=b$:

condiciones:

1) $x^{(0)}$ punto inicial debe ser factible: $x^{(0)}\in$ dom$f$ y $Ax^{(0)}=b$
2) El paso de Newton, Δx_{nt}, debe modificarse de modo que satisfaga las restricciones

Para esto la idea consiste en:

$$\begin{array}{c}\min f(x)\\ \text{s.a. } Ax=b\end{array} \rightsquigarrow \begin{array}{c}\min \hat{f}(x+v)\\ \text{s.a. } A(x+v)=b\end{array} \quad \text{con } \hat{f}(x+v)=f(x)+\nabla f(x)^T v+\frac{1}{2}v^T\nabla^2 f(x)v$$

y como f es convexa este problema es de minimización cuadrática con restricciones de igualdad y tiene solución $(\Delta x_{nt}, w)$ que satisfacen al sistema: $\begin{bmatrix}\nabla^2 f(x) & A^T\\ A & 0\end{bmatrix}\begin{bmatrix}\Delta x_{nt}\\ w\end{bmatrix}=\begin{bmatrix}-\nabla f(x)\\ 0\end{bmatrix}$ el cual resulta de las condiciones de KKT y del hecho que como $Ax=b$ y $A(x+v)=b$ entonces $Av=0$ | en este caso $v=\Delta x_{nt}$ y w es variable dual asociada

$\hookrightarrow$ $\min \; x^T P x + q^T x + r$ $\quad P \in S_+^n, \; A \in \mathbb{R}^{p\times n}$ entonces:
s.a. $\; Ax = b$

cond KKT: $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$

Obs

•) El paso de Newton Δx_{nt} está definido para puntos en los que la matriz KKT es no singular

•) $x + \Delta x_{nt}$ estimación de x^*, w estimación de ν^*

•) Decremento de Newton: $\lambda(x) = \|\Delta x_{nt}\|_{\nabla^2 f(x)} = \left(\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt}\right)^{1/2}$

$\hookrightarrow$ Ayuda al backtracking pues: $\nabla f(x)^T \Delta x_{nt} = -\lambda^2(x)$

$\hookrightarrow$ Ayuda al criterio de paro: $\frac{\lambda^2(x)}{2} <$ tolerancia, $\lambda^2(x)$ estima $f(x) - p^*$

Método de Newton con restricciones de igualdad

Dados un punto inicial $x \in \operatorname{dom} f$ con $Ax = b$ y una tolerancia $\varepsilon > 0$.

Repetir:
1. Calcular el paso y decremento de Newton Δx_{nt}, $\lambda(x)$.
2. Criterio de paro: terminar método si $\frac{\lambda^2(x)}{2} \leq \varepsilon$.
3. Búsqueda de línea: Elegir tamaño de paso por método de backtracking line search.
4. Actualizar: $x + t\Delta x_{nt}$.