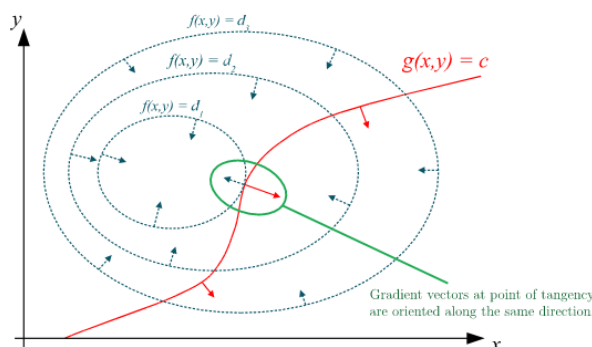


## Dualidade

Fonte: [http://stanford.edu/~boyd/cvxbook/bv\\_cvxbook.pdf](http://stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf)

A técnica do multiplicador de Lagrange permite encontrar o máximo ou mínimo de uma função  $f(x_1, \dots, x_n)$  quando há restrições nos valores das variáveis de entrada, ou seja se aplica a restrições do tipo  $g(x_1, \dots, x_n) = c$ . A restrição  $g$  é uma função com o mesmo domínio de  $f$ , e  $c$  é uma constante.

A principal ideia é procurar pontos onde as curvas de nível de  $f$  e  $g$  sejam tangentes entre si. Isto é o mesmo que encontrar pontos onde os vetores gradientes de  $f$  e  $g$  são paralelos entre si. Quando as curvas de nível de duas funções  $f$  e  $g$  são tangentes, seus gradientes são paralelos.



O fato das curvas de nível serem tangentes não diz nada sobre a magnitude de cada um dos gradientes. Quando dois vetores apontam na mesma direção significa que podemos multiplicar um por alguma constante para obter o outro. Especificamente, seja  $(x_0, y_0)$  um ponto em que as linhas de  $f$  e  $g$  são tangentes. Como essa tangência significa que seus vetores de gradiente estão alinhados, podemos escrever:

$$\nabla f(x_0, y_0) = \lambda_0 \nabla g(x_0, y_0), \lambda_0 \in \mathbb{R}$$

Como temos a comparação de 2 vetores, podemos “quebrar” a equação em duas.

$$\begin{cases} f_x(x_0, y_0) = \lambda_0 g_x(x_0, y_0) \\ f_y(x_0, y_0) = \lambda_0 g_y(x_0, y_0) \end{cases}$$

Nos anos 1700, Joseph Louis **Lagrange** escreveu uma nova função especial que engloba todas as variáveis de entrada de  $f$  e  $g$  e a nova variável  $\lambda$ .

$$\mathcal{L}(x_0, y_0, \lambda) = f(x_0, y_0) + \lambda(g(x_0, y_0) - c)$$

As três condições que precisamos resolver para encontrar  $x$ ,  $y$ , e  $\lambda$  passam por todas as derivadas parciais de  $\mathcal{L}$  sendo iguais a 0.

$$\begin{cases} f_x(x_0, y_0) - \lambda g_x(x_0, y_0) = 0 \\ f_y(x_0, y_0) - \lambda g_y(x_0, y_0) = 0 \\ g(x_0, y_0) - c = 0 \end{cases}$$

Isso pode ser escrito de forma compacta impondo o gradiente de  $\mathcal{L}$  igual ao vetor zero:

$$\nabla \mathcal{L} = 0$$

A função  $\mathcal{L}$  é denominado de Lagrangeano, e a nova variável  $\lambda$  de multiplicador de Lagrange.

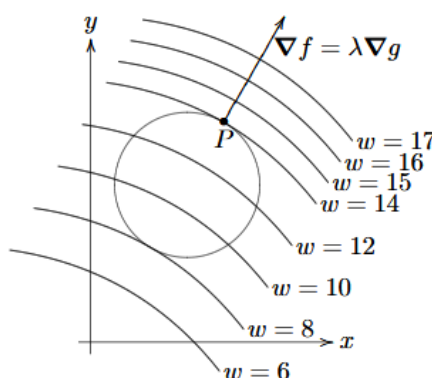
## Teorema de Lagrange

Sejam  $f$  e  $g$  funções de classe  $C^1$ , definidas num domínio aberto comum  $D \subseteq \mathbb{R}^2$ . Seja  $C = \{(x, y) \in D : g(x, y) = c\}$  um conjunto não vazio. Se  $(x_0, y_0) \in C$  é um ponto extremo de  $f$  em  $C$  e  $\nabla g(x_0, y_0) \neq 0$ , então existe um número  $\lambda \in \mathbb{R}$ , tal que  $\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$ .

Esse teorema nos dá uma condição necessária (mas não suficiente) para que um ponto  $(x_0, y_0) \in C$  seja um ponto extremo de  $f$ .

### Explicação/ilustração 1 do Teorema

Consideramos apenas o caso bidimensional,  $w = f(x, y)$  com restrição  $g(x, y) = c$ . Para um melhor entendimento, desenhamos a curva de restrição,  $g(x, y) = c$ , que é um círculo, e algumas curvas de nível para  $w = f(x, y) = k$  com valores explícitos ("chutados"). Geometricamente, estamos procurando por pontos do círculo onde  $w$  assume seus valores máximos ou mínimos.



<http://bit.do/eRpPf>

Começamos na curva de nível com  $w=17$ , que não tem pontos no círculo. Então, claramente, o valor máximo de  $w$  no círculo de restrição é menor que 17. Mova as curvas de nível até tocarem o círculo quando  $w=14$ . Denominemos o ponto em que ocorre o primeiro toque de  $P$ . É claro que  $P$  dá um máximo local para  $w$  em  $g=c$ , porque se você se afastar de  $P$  em qualquer direção no círculo, você estará em uma curva de nível com um valor menor.

Como o círculo é uma curva de nível para  $g$ , sabemos que o  $\nabla g$  é perpendicular a ele. Sabemos também que  $\nabla f$  é perpendicular à curva de nível  $w=14$ , uma vez que as próprias curvas são tangentes, estes dois gradientes devem ser paralelos. Da mesma forma, se você continuar descendo as curvas de nível, a última a tocar o círculo dará um mínimo local e o mesmo argumento será aplicado. ■

### Explicação/ilustração 2 do Teorema

Suponha que  $f(x)$  é uma função de várias variáveis, ou seja,  $x$  é um vetor e  $g(x) = c$  é uma restrição. Então o teorema do multiplicador de Lagrange diz que no ponto máximo de  $f$  sujeito à restrição  $g$  temos  $\nabla f = \lambda \nabla g$ .

De onde vem  $\lambda$ ? E por que o gradiente de sua função objetivo deveria estar relacionado ao gradiente de uma restrição? Estas parecem ser duas coisas diferentes que nem deveriam ser comparáveis.

O conjunto de pontos que satisfazem  $g(x) = c$  é uma superfície. E para qualquer  $k$ , o conjunto de pontos que satisfazem  $f(x) = k$  também é uma superfície. Imagine  $k$  muito grande, maior que o máximo de  $f$  na superfície definida por  $g(x) = c$ . Você pode pensar na superfície  $g(x) = c$  sendo um balão dentro do balão maior  $f(x) = k$ .

Agora diminua gradualmente  $k$ , como deixar o ar sair do balão externo, até o primeiro toque das superfícies  $g(x) = c$  e  $f(x) = k$ . Nesse ponto, as duas superfícies serão tangentes e, assim, seus vetores normais, dados por seus gradientes, apontarão na mesma direção. Ou seja,  $f$  e  $\nabla g$  são paralelos e, portanto,  $\nabla f$  é múltiplo de  $\nabla g$ . Chamaremos este múltiplo de  $\lambda$ . ■

### Função Lagrangeana – Restrições de igualdade

Portanto, quando maximizar/minimizar uma função  $f(x_1, \dots, x_n)$  sujeita à restrição de uma função  $g(x_1, \dots, x_n) = c$ , devemos introduzir a nova variável  $\lambda$  e definir a função Lagrangeana  $\mathcal{L}$ :

$$\mathcal{L}(x_1, \dots, x_n, \lambda) = f(x_1, \dots, x_n) + \lambda(g(x_1, \dots, x_n) - c)$$

Para determinar os pontos críticos da função  $\mathcal{L}$  basta resolver

$$\nabla \mathcal{L}(x_1, \dots, x_n, \lambda) = 0$$

Se considerarmos um problema da forma maximizar/minimizar  $f(x_1, \dots, x_n)$  sujeito às duas restrições  $g_1(x_1, \dots, x_n) = c_1$  e  $g_2(x_1, \dots, x_n) = c_2$ , definimos a função Lagrangeana  $\mathcal{L}$  como

$$\mathcal{L}(x_1, \dots, x_n, \lambda_1, \lambda_2) = f(x_1, \dots, x_n) + \lambda_1(g_1(x_1, \dots, x_n) - c_1) + \lambda_2(g_2(x_1, \dots, x_n) - c_2)$$

sendo necessário resolver

$$\nabla \mathcal{L}(x_1, \dots, x_n, \lambda_1, \lambda_2) = 0$$

A resolução do sistema, composto por  $(n + 2)$  equações e  $(n + 2)$  incógnitas, fornece todos os possíveis candidatos para os máximos e mínimos locais.

E para o caso geral com  $n$  variáveis e  $m$  restrições, o problema é dado

$$\text{maximize (ou minimize) } f(x), \text{ sujeito a } g_j(x) = 0, j = 1, 2, \dots, m, \text{ onde } x = (x_1, \dots, x_n)^T$$

Neste caso, a função de Lagrangeana,  $\mathcal{L}$ , terá um multiplicador de Lagrange  $\lambda_j$  para cada restrição:

$$\mathcal{L}(x, \lambda_1, \lambda_2, \dots, \lambda_m) = f(x) + \sum_{j=1}^m \lambda_j g_j(x)$$

As condições necessárias são

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f}{\partial x_i}(x) + \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i}(x), i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \lambda_j} = g_j(x), j = 1, \dots, m \end{cases}$$

formada por  $(n + m)$  equações em função de  $(n + m)$  incógnitas.

### Função Lagrangeana – Restrições de desigualdade

Fonte 1: <https://people.eecs.berkeley.edu/~klein/papers/lagrange-multipliers.pdf>

Fonte 2: [http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/306532/1/Padua\\_SuzanGrazielleBenettide\\_M.pdf](http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/306532/1/Padua_SuzanGrazielleBenettide_M.pdf)

O método do multiplicador de Lagrange também cobre o caso de restrições de desigualdade do tipo  $g(x) \leq b$ . Podemos resolver problemas do tipo

$$\text{maximize } f(x), \text{ sujeito a } h(x) \leq b, \text{ com } x = (x_1, \dots, x_n)^T$$

Caso  $P$  seja solução do problema de otimização, isto é,  $P$  é o ponto que maximiza  $f$  no conjunto viável  $D$ , temos duas possibilidades:  $P$  está na fronteira do conjunto viável  $D$  ( $h(P) = b$ ), ou,  $P$  está no interior do conjunto viável  $D$  ( $h(P) < b$ ).

- (i) Se  $h(P) = b$ , dizemos que a restrição  $g$  está ativa no ponto  $P$ . Geometricamente, significa dizer que o ponto  $P$  está na fronteira do conjunto viável  $D$ . Então, estamos em um caso parecido com a otimização com uma restrição de igualdade, valendo a ideia do Teorema dos Multiplicadores de Lagrange de que se  $P$  é um ponto de máximo de  $f$  em  $D$  então os gradientes de  $f$  e  $h$  são paralelos nesse ponto. Logo, existe um número real  $\mu^*$  tal que

$$\nabla f(P) = \mu^* \nabla h(P)$$

Como o ponto de máximo está na fronteira, e não no interior de  $D$ , temos que o gradiente de  $f$  no ponto  $P$  tem que apontar para “fora” da região  $D$ , afinal, quando não nulo, ele fornece a direção de maior crescimento da  $f$  no ponto. O gradiente de  $h$  no ponto  $P$  também fornece a direção de maior crescimento desta função no ponto. Como dentro do conjunto viável temos que  $h(P) < b$  e na fronteira temos  $h(P) = b$ , logo  $\nabla h(P)$  também aponta para fora da região  $D$ . Assim,  $\nabla h(P)$  e  $\nabla f(P)$ , além de serem paralelos, têm o mesmo sentido: os dois apontam para “fora” da região viável, o que implica que  $\mu^*$  é um número não negativo,

$$\mu^* \geq 0$$

Desta forma temos

$$\nabla f(P) = \mu^* \nabla h(P), \mu^* \geq 0, h(P) = b$$

- (ii) Se  $h(P) < b$ , dizemos que a restrição  $h$  não está ativa em  $P$ . Geometricamente, significa dizer que o ponto  $P$  está no interior do conjunto viável  $D$ . Então estamos em um caso de otimização sem restrição e podemos aplicar a regra de Fermat, pois  $P$  é um ponto crítico de  $f$  no interior do conjunto viável  $D$ , ou seja,

$$\nabla f(P) = 0$$

Desta forma

$$\nabla f(P) = 0, h(P) < b$$

Podemos unificar as possibilidades (i) e (ii) através da equação  $\mu^*[h(P) - b] = 0$ , denominada **condição de complementaridade**. As soluções dessa equação são  $\mu^* = 0$  ou  $h(P) = b$ .

- a) Se  $\mu^* = 0$  temos  $\nabla f(P) = \mu^* \nabla h(P) = 0 \times \nabla h(P) = 0$ , isto é,  $P$  é um ponto crítico de  $f$  (podendo ocorrer  $h(P) < b$  ou  $h(P) = b$ ). Se  $h(P) < b$  então a restrição  $h$  não está ativa no ponto  $P$ . Se, porém,  $h(P) = b$  então temos o caso particular em que  $P$  é um ponto crítico de  $f$  na fronteira do conjunto viável  $D$ .
- b) Por outro lado, se  $\mu^* > 0$  temos  $h(P) = b$ , isto é, a restrição  $h$  está ativa em  $P$  e, portanto vale  $\nabla f(P) = \mu^* \nabla h(P)$ , com  $\mu^* > 0$ .

Portanto, se a restrição for do tipo  $\leq$  num problema de maximização, então o multiplicador de Lagrange é positivo, ou seja  $\mu \geq 0$ .

**Observação:** Um aspecto útil do método do multiplicador de Lagrange é que os valores dos multiplicadores nos pontos de solução geralmente têm algum significado. Matematicamente, um multiplicador  $\lambda$  é o valor da derivada parcial de  $\mathcal{L}$  em relação à restrição  $g_i$ . Portanto, é a taxa em que poderíamos aumentar o Lagrangiano se elevássemos “a meta” dessa restrição (de zero). Mas lembre-se que nos pontos de solução  $x$ ,  $\mathcal{L}(x, \lambda) = f(x)$ . Portanto, a taxa de aumento do Lagrangiano em relação a essa restrição é também a taxa de aumento do máximo valor restrito de  $f$  com relação a essa restrição. Em economia, quando  $f$  é uma função de lucro e o  $g_i$  é uma restrição de recursos,  $\lambda_i$  seria a quantia (possivelmente negativa!) pela qual o lucro subiria se fosse permitido mais uma unidade do recurso  $i$ . Essa taxa é chamada de **preço-sombra** de  $i$ , que é

interpretada como a quantia que valeria a pena para relaxar a restrição (por P & D, mineração, suborno ou qualquer outro meio). (Fonte: <http://www.cs.cmu.edu/~ggordon/lp.pdf>)

### Teorema dos multiplicadores de Lagrange – Equações lineares

<http://www.cs.cmu.edu/~ggordon/lp.pdf>

#### A – PL: minimização, $Ax \leq b$ e $x$ livre

Considere o pl

$$\begin{array}{ll} \min & f(x) = c^T x \\ \text{sa} & Ax \leq b \end{array} \quad \Rightarrow \quad \begin{array}{ll} \min & f(x) = c^T x \\ \text{sa} & Ax - b \leq 0 \end{array}$$

Vamos nos referir ao pl acima como o problema **primal**, e à variável de decisão  $x$  nesse problema, como a variável primais. Um propósito da dualidade (de Lagrange) é encontrar um limite inferior em um problema de minimização (ou limites superiores para um problema de maximização). Para isso, precisamos construir o Lagrangeano.

Para formar o Lagrangiano introduzimos multiplicadores de Lagrange  $\mu$  e obtemos

$$\mathcal{L}(x, \mu) = c^T x + \mu^T (Ax - b) = -b^T \mu + (A^T \mu + c)^T x$$

Observamos que, para todo possível  $x \in X$ , e todo  $\mu \geq 0$ ,  $f(x)$  é limitado abaixo por  $\mathcal{L}(x, \mu)$ , ou seja

$$\forall x \in X, \forall \mu \in \mathbb{R}_+^m : f(x) \geq \mathcal{L}(x, \mu)$$

A função Lagrange dual, ou simplesmente a função dual é:

$$g(\mu) = \min_x \mathcal{L}(x, \mu) = \min_x \left( -b^T \mu + (A^T \mu + c)^T x \right) = -b^T \mu + \min_x (A^T \mu + c)^T x$$

Temos que  $g(\mu) = +\infty$  exceto quando  $A^T \mu + c = 0$ , caso em que  $g(\mu) = -b^T \mu$ , ou seja:

$$g(\mu) = \min_x \mathcal{L}(x, \mu) = \begin{cases} -b^T \mu, & A^T \mu + c = 0 \\ +\infty, & \text{caso contrário} \end{cases}$$

O problema Lagrange dual do pl acima é maximizar a função dual em relação a  $\mu \geq 0$ , ou seja,

$$\max_{\mu} g(\mu) = \max_{\mu} \{-b^T \mu : A^T \mu + c = 0, \mu \geq 0\}$$

Ou seja, o pl dual é:

$$\begin{array}{ll} \max & -b^T \mu \\ \text{sa} & A^T \mu + c = 0 \\ & \mu \geq 0 \end{array}$$

#### B – PL: minimização, $Ax = b$ e $x \geq 0$ (pl na forma padrão)

Considere o pl

$$\begin{array}{ll} \min Z = & c^T x \\ \text{sa} & Ax = b \\ & x \geq 0 \end{array} \quad \Rightarrow \quad \begin{array}{ll} \min Z = & c^T x \\ \text{sa} & Ax - b = 0 \\ & -x \leq 0 \end{array}$$

Para formar o Lagrangiano introduzimos multiplicadores de Lagrange  $\lambda_i$  para as  $m$  restrições de igualdade e  $\mu_j$  para as  $n$  desigualdades e obtemos

$$\mathcal{L}(x, \lambda, \mu) = c^T x + \lambda^T (b - Ax) - \mu^T x = -b^T \lambda + (c - A^T \lambda - \mu)^T x$$

Se  $\mu_i < 0$  para algum  $i$  então  $-\mu^T x \rightarrow +\infty$ . Portanto só existe um máximo finito  $x \geq 0$  se  $\mu_i \geq 0 \forall i$ .

A função dual é

$$g(\lambda, \mu) = \min_x \mathcal{L}(x, \lambda, \mu) = \min_x (b^T \lambda + (c - A^T \lambda - \mu)^T x) = b^T \lambda + \min_x (c - A^T \lambda - \mu)^T x$$

Temos  $g(\lambda, \mu) = \infty$  exceto quando  $c = A^T \lambda + \mu$ , caso em que é  $g(\lambda, \mu) = b^T \lambda$ , ou seja:

$$g(\lambda, \mu) = \begin{cases} b^T \lambda, & c = A^T \lambda + \mu \\ +\infty, & \text{caso contrário} \end{cases}$$

O problema Lagrange dual de um pl na forma padrão é maximizar a função dual em relação a  $\mu \geq 0$ , ou seja,

$$\begin{aligned} \max \quad & g(\lambda, \mu) = \begin{cases} b^T \lambda, & c = A^T \lambda + \mu \\ \infty, & \text{caso contrário} \end{cases} \\ \text{sa} \quad & \mu \geq 0 \end{aligned}$$

Aqui  $g$  é finito somente quando  $c = A^T \lambda + \mu$ . Podemos formar o pl dual:

$$\begin{aligned} \max \quad & b^T \lambda \\ \text{sa} \quad & A^T \lambda + \mu = c \\ & \mu \geq 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} \max \quad & b^T \lambda \\ \text{sa} \quad & A^T \lambda \leq c \\ & \lambda \text{ livre} \end{aligned}$$

## B – PL: minimização, $Ax = b$ e $x \geq 0$ (sem utilizar $x \geq 0$ no Lagrangeano)

(Fonte: <http://www.fc.unesp.br/~adriana/Pos/PO7e8.pdf>)

Considere o pl

$$\begin{aligned} \min \quad & f(x) = c^T x \\ \text{sa} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Como vimos anteriormente, não é necessário utilizar  $x \geq 0$  no Lagrangeano.

$$\mathcal{L}(x, \lambda) = c^T x + \lambda^T (b - Ax) = (c^T - \lambda^T A)x + \lambda^T b$$

Seja  $A = [a_1, a_2, \dots, a_n]$ ,  $c = (c_1, c_2, \dots, c_n)$ ,  $(c^T - \lambda^T A) = (c_1 - \lambda^T a_1, c_2 - \lambda^T a_2, \dots, c_n - \lambda^T a_n)$ , então

$$\mathcal{L}(x, \lambda) = c^T x + \lambda^T (b - Ax) = (c_1 - \lambda^T a_1)x_1 + (c_2 - \lambda^T a_2)x_2 + \dots + (c_n - \lambda^T a_n)x_n + \lambda^T b$$

com domínio  $\mathcal{L} = \mathbb{R}_+^n \times \mathbb{R}^m$ . A função Lagrange dual, ou simplesmente a função dual é

$$g(\lambda) = \min_{x \geq 0} \mathcal{L}(x, \lambda) = \min_{x \geq 0} [(c^T - \lambda^T A)x + \lambda^T b] = \min_{x \geq 0} [(c^T - \lambda^T A)x] + \lambda^T b$$

$$g(\lambda) = \min_{x \geq 0} [(c_1 - \lambda^T a_1)x_1 + (c_2 - \lambda^T a_2)x_2 + \dots + (c_n - \lambda^T a_n)x_n] + \lambda^T b$$

$$g(\lambda) = \min_{x_1 \geq 0} (c_1 - \lambda^T a_1)x_1 + \min_{x_2 \geq 0} (c_2 - \lambda^T a_2)x_2 + \dots + \min_{x_n \geq 0} (c_n - \lambda^T a_n)x_n + \lambda^T b$$

Se algum coeficiente de alguma variável  $x$  for negativo, então  $g(\lambda) = -\infty$ . Se o coeficiente da variável  $x_i$  for nulo, então  $x_i$  pode assumir qualquer valor não negativo sem que a função dual  $g(\lambda)$  se altere.

Assim, a primeira parte de  $\mathcal{L}(x, \lambda)$ , que depende de  $x_1, \dots, x_n$ , é sempre  $-\infty$  ou zero. Se for zero, então  $\mathcal{L}(x, \lambda) = \lambda^T b$ . A definição da função dual fornece uma desigualdade fundamental da teoria da dualidade: limitantes inferiores para os problemas de minimização (superiores para problemas de maximização).

**Teorema:**  $g(\lambda) \leq f(x)$ ,  $\forall \lambda \in \mathbb{R}^m$  e  $\forall x$  tal que  $Ax = b$ ,  $x \geq 0$ .

Seja  $R = \{x \in \mathbb{R}^m | x \geq 0\} \supseteq S = \{x \in \mathbb{R}^m | Ax = b, x \geq 0\}$ . Então

$$\begin{aligned} g(\lambda) &= \min_{x \geq 0} \mathcal{L}(x, \lambda) = \min_{x \geq 0} [c^T x + \lambda^T (b - Ax)] \\ &\leq \min_{Ax=b, x \geq 0} [c^T x + \lambda^T (b - Ax)] \quad (\text{lembreemos que } Ax = b) \\ &= \min_{\substack{Ax=b \\ x \geq 0}} f(x) \\ &\leq f(x), \forall x \in S \end{aligned}$$

O maior limitante inferior para  $f(x)$ , obtido pela função dual, define o problema dual:

$$\max_{\lambda} g(\lambda), \lambda \in \mathbb{R}^m$$

onde  $\lambda_1, \lambda_2, \dots, \lambda_m$  são denominados de variáveis duais.

Para cada escolha das variáveis duais, o problema lagrangeano é facilmente resolvido como uma soma de  $n$  subproblemas.

$$\begin{aligned} g(\lambda) &= \min_{x \geq 0} \mathcal{L}(x, \lambda) = \min_{x \geq 0} [(c^T - \lambda^T A)x + \lambda^T b] \\ &= \min_{x \geq 0} [(c_1 - \lambda^T a_1)x_1 + (c_2 - \lambda^T a_2)x_2 + \dots + (c_n - \lambda^T a_n)x_n] + b^T \lambda \\ &= \min_{x_1 \geq 0} (c_1 - \lambda^T a_1)x_1 + \min_{x_2 \geq 0} (c_2 - \lambda^T a_2)x_2 + \dots + \min_{x_n \geq 0} (c_n - \lambda^T a_n)x_n + b^T \lambda \end{aligned}$$

Seja o subproblema associado a determinação de  $x_j$ , então

$$\min_{x_j \geq 0} (c_j - \lambda^T a_j)x_j \Rightarrow \begin{cases} x_j = 0, & \text{se } c_j - \lambda^T a_j \geq 0 \\ x_j = -\infty, & \text{se } c_j - \lambda^T a_j < 0 \end{cases}$$

Se  $c_j - \lambda^T a_j < 0$ , então  $g(\lambda) = -\infty$  (que pode ser considerado um limitante inferior). Assim, devemos escolher  $\lambda$  de tal forma que

$$c_j - \lambda^T a_j \geq 0 \Rightarrow g(\lambda) = b^T \lambda$$

Assim,

$$\lambda^T a_1 \leq c_1, \lambda^T a_2 \leq c_2, \dots, \lambda^T a_n \leq c_n \Leftrightarrow \lambda^T A \leq c^T \Leftrightarrow A^T \lambda \leq c$$

Então o dual é,

$$\begin{aligned} \max \quad & \lambda^T b \\ \text{sa} \quad & A^T \lambda \leq c \\ & \lambda \text{ livre} \end{aligned}$$

## C – PL: maximização, $Ax \leq b$ e $x \geq 0$ (Fonte: <http://www.statslab.cam.ac.uk/~rrw1/opt/O.pdf>)

Considere o pl

$$\begin{aligned} \max \quad & c^T x \\ \text{sa} \quad & Ax \leq b \\ & x \geq 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} \max \quad & c^T x \\ \text{sa} \quad & Ax + F - b = 0 \\ & x, F \geq 0 \end{aligned}$$

Desconsiderando  $x \geq 0$  na função Lagrangeana, temos

$$\mathcal{L}(x, F, \mu) = c^T x - \mu^T (Ax + F - b) = (c^T - \mu^T A)x - \mu^T F + \mu^T b$$

Podemos encontrar o conjunto  $Y$  tal que  $\mu \in Y$  implica em  $\max_{x, F \geq 0} \{\mathcal{L}(x, F, \mu)\}$  finito, e para  $\mu \in Y$  calculamos o mínimo de  $g(\mu)$ . Considere o termo linear  $-\mu^T F$ . Se qualquer coordenada  $\mu_i < 0$ , então podemos tornar  $-\mu_i F_i$  grande o quanto quisermos, tomando  $F_i$  grande. Portanto, existe um máximo finito em  $F \geq 0$  se  $\mu_i \geq 0 \forall i$ .

Da mesma forma, o termo  $(c^T - \mu^T A)x$  pode ser tornado tão grande quanto quisermos, a menos que  $(c_i^T - \mu^T a_i) \leq 0$  para todo  $i$ . Assim,

$$Y = \{\mu : \mu \geq 0, c^T - \mu A \leq 0\} \Rightarrow Y = \{\mu : \mu \geq 0, \mu A \geq c^T\}.$$

Se escolhermos um  $\mu \in Y$  então

$$\max_{F \geq 0} \{-\mu^T F\} = 0$$

escolhendo  $F_i = 0$  se  $\mu_i > 0$  e qualquer  $F_i$  se  $\mu_i = 0$  e também

$$\max_{x \geq 0} (c^T - \mu^T A)x = 0$$

similarmente. Portanto, para  $\mu \in Y$ ,  $g(\mu) = \mu^T b$ .

Então o dual é,

$$\begin{aligned} \min \quad & \mu^T b \\ \text{sa} \quad & A^T \mu \geq c^T \\ & \mu \geq 0 \end{aligned}$$

A intuição chave por trás da teoria da dualidade de Lagrange é a seguinte: Para qualquer problema de otimização convexa, sempre existem configurações das variáveis duais, de modo que o mínimo irrestrito do Lagrangeano em relação às variáveis primais (mantendo as variáveis duais fixas) coincida com a solução do problema original de minimização restrita.