UNIVERSIDADE
DE VIGO

Departamento de Estatística
e Investigación Operativa

Minería de Datos, 2008-2009
Optimización con restricciones

Problema de Optimización con Restricciones

Lo que sigue puede verse de forma detallada, p. ej. en (Chong y Zak 2001) y (Luenberger 1984).

Definición 1 *Un problema (**Primal**) de optimización con restricciones se formula de la siguiente forma: dadas funciones* $f, \mathbf{g} = (g_1, ..., g_r), \mathbf{h} = (h_1, ..., h_m)^T$*, con* f, g_i, h_j*,* $i = 1, ..., r$*,* $j = 1, ..., m$*, definidas en* $\Omega \subseteq \mathbb{R}^p$*, determinar:*

$$\begin{cases} \min\limits_{\mathbf{w}\in\Omega} f(\mathbf{w}) \text{ sujeto a:} \\ \qquad \mathbf{g}(\mathbf{w}) \leq \mathbf{0} \\ \qquad \mathbf{h}(\mathbf{w}) = \mathbf{0} \end{cases} \tag{1}$$

Definición 2 *La función* $f(\mathbf{w})$ *se denomina **función objetivo o de coste**. Se denomina **región factible** la región en la que el problema está definido:* $R = \{\mathbf{w} \in \Omega : \mathbf{g}(\mathbf{w}) \leq \mathbf{0}, \mathbf{h}(\mathbf{w}) = \mathbf{0}\}$*. Una restricción de desigualdad* $g_i(\mathbf{w}) \leq 0$ *se dice **activa** si, en el óptimo* $\mathbf{w}^*$*, supuesto existente, verifica:* $g_i(\mathbf{w}^*) = 0$*. Se denomina **slack**, a una variable* $\xi_i \geq 0$ *asociada a la restricción* $g_i(\mathbf{w}) \leq 0$ *que la convierte en activa:* $g_i(\mathbf{w}) + \xi_i = 0$*. Si la restricción es activa,* $\xi_i = 0$*.*

Definición 3 *Una función* $f : \mathbb{R}^p \to \mathbb{R}$ *se dice **convexa** si* $\forall \mathbf{u}, \mathbf{w} \in \mathbb{R}^p$ *y* $\forall \lambda \in (0, 1)$:

$$f(\lambda\mathbf{u} + (1-\lambda)\mathbf{w}) \leq \lambda f(\mathbf{u}) + (1-\lambda)f(\mathbf{w})$$

*y **estrictamente convexa** si la desigualdad es estricta. Un conjunto* $\Omega \subseteq \mathbb{R}^p$ *se dice **convexo** si* $\forall \mathbf{u}, \mathbf{w} \in \Omega$ *y* $\forall \lambda \in (0, 1) : \lambda\mathbf{u} + (1-\lambda)\mathbf{w} \in \Omega$*.*

Lema 4 *Si una función* f *es convexa, cualquier mínimo local* $\mathbf{w}^*$ *del problema sin restricciones,* $\min_{\mathbf{w}\in\Omega} f(\mathbf{w})$*, es también mínimo global.*

Demostración. $\forall \mathbf{u} \neq \mathbf{w}^*$, $\exists \lambda \in (0, 1)$, tal que:

$$f(\mathbf{w}^*) \leq f(\lambda\mathbf{w}^* + (1-\lambda)\mathbf{u}) \leq \lambda f(\mathbf{w}^*) + (1-\lambda)f(\mathbf{u}) \Rightarrow f(\mathbf{w}^*) \leq f(\mathbf{u})$$

■

Definición 5 *Un problema de optimización con restricciones se dice **convexo** si* Ω*, la función objetivo y todas las restricciones son convexas.*

Teorema 6 ***Fermat**. Una condición necesaria para que* $\mathbf{w}^*$ *sea mínimo de* $f(\mathbf{w})$ *(sin restricciones), con* f *función diferenciable en* $\mathbf{w}^*$*, es que* $\nabla_{\mathbf{w}} f(\mathbf{w}^*) = \mathbf{0}$*. Si, además,* f *es convexa, dicha condición es suficiente.*

Definición 7 *Dado un problema de optimización con función objetivo $f(\mathbf{w})$ y restricciones de igualdad $\mathbf{h}(\mathbf{w}) = \mathbf{0}$, se define la* ***Función de Lagrange (Lagrangiano)*** *como la función:*

$$L(\mathbf{w}, \lambda) = f(\mathbf{w}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{w})$$

donde los λ_i se denominan ***coeficientes de Lagrange****.*

Teorema 8 ***Lagrange****. Una condición necesaria para que $\mathbf{w}^*$ sea solución de $\min_{\mathbf{w}\in\Omega} f(\mathbf{w})$ sujeto a $\mathbf{h}(\mathbf{w}) = \mathbf{0}$, con $f, \mathbf{h}$ diferenciables en un entorno de $\mathbf{w}^*$ es que, exista $\boldsymbol{\lambda}^* = (\lambda_1^*, ..., \lambda_m^*)$ tal que:*

$$\begin{cases} \nabla_{\mathbf{w}} f(\mathbf{w}^*, \boldsymbol{\lambda}^*) = \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} f(\mathbf{w}^*, \boldsymbol{\lambda}^*) = \mathbf{0} \end{cases}$$

Si, además, $L(\mathbf{w}, \boldsymbol{\lambda}^*)$ es convexa como función de $\mathbf{w}$, entonces dichas condiciones son suficientes.

Los coeficientes de Lagrange expresan la tasa de variación del valor óptimo $f(\mathbf{w}^*)$ ante cambios en el valor de las restricciones, pues:

$$\frac{\partial L(\mathbf{w}^*, \boldsymbol{\lambda}^*)}{\partial h_i} = \lambda_i^*,\ i = 1 : m$$

Definición 9 *Dado un problema de optimización con restricciones de igualdad y desigualdad, (1), se define la* ***función de Lagrange (Lagrangiano) generalizada*** *como la función, con $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_r)$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_m)$:*

$$L(w, \alpha, \beta) = f(w) + \sum_{i=1}^{r} \alpha_i g_i(w) + \sum_{i=1}^{m} \beta_i h_i(w)$$

Definición 10 *Se define el problema de optimización* ***Dual****, asociado al problema* ***(Primal)*** *(1), como el problema:*

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left\{ \inf_{\mathbf{w}\in\Omega} L(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \right\}$$
$$\textit{sujeto a: } \boldsymbol{\alpha} \geq 0$$

donde $\rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \inf_{\mathbf{w}\in\Omega} L(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Si $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ es la solución óptima, se denomina ***valor del problema*** *a: $\rho(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$.*

Teorema 11 *Sea $\mathbf{w} \in \Omega$ una solución factible del problema primal y $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ solución factible del problema dual. Entonces: $f(\mathbf{w}) \geq \rho(\boldsymbol{\alpha}, \boldsymbol{\beta})$.*

Demostración. Se verifica que:

$$\rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq \inf_{\mathbf{u}\in\Omega} L(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \leq L(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f(\mathbf{w}) + \boldsymbol{\alpha}^T \mathbf{g}(\mathbf{w}) + \boldsymbol{\beta}^T \mathbf{h}(\mathbf{w}) \leq f(\mathbf{w})$$

pues, al ser $\mathbf{w}$ factible, $\mathbf{g}(\mathbf{w}) \leq \mathbf{0}$, $\mathbf{h}(\mathbf{w}) = \mathbf{0}$ y al ser $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ factible se tiene $\boldsymbol{\alpha} \geq \mathbf{0}$. ■

De esta forma, el valor del problema dual es siempre menor o igual que el valor del problema primal.

Corolario 12 *Si se verifica que* $f(\mathbf{w}^*) = \rho(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$, *siendo* $\mathbf{w}^*$ *y* $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ *soluciones factibles para los problemas primal y dual respectivamente, entonces* $\mathbf{w}^*$ *y* $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ *son soluciones del problema primal y dual, respectivamente, en cuyo caso:* $\alpha_i^* g_i(\mathbf{w}^*) = 0$, $i = 1 : r$.

Demostración. La condición de óptimos para los respectivos problemas está clara pues alcanzan sus cotas inferior y superior. Además, como en la demostración del teorema anterior, por hipótesis, las desigualdades son igualdades, ha de ser $\boldsymbol{\alpha}^{*T}\mathbf{g}(\mathbf{w}^*) = 0$. ■

Los problemas primal y dual pueden poseer soluciones óptimas sin que se alcance la igualdad anterior, en cuyo caso, se denomina **gap de dualidad** a la diferencia $f(\mathbf{w}^*) - \rho(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$.

Definición 13 *Un punto* $(\mathbf{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ *se dice* ***Punto de Silla*** *para el Lagrangiano generalizado si* $\mathbf{w}^* \in \Omega, \boldsymbol{\alpha}^* \geq 0$ *y:*

$$L(\mathbf{w}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}) \leq L(\mathbf{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \leq L(\mathbf{w}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$$

Teorema 14 $(\mathbf{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ *es punto de silla del Lagrangiano generalizado si y sólo si* $\mathbf{w}^*$ *y* $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ *son soluciones del problema primal y dual, respectivamente y no existe gap de dualidad.*

Teorema 15 ***Teorema de Dualidad Fuerte****. Si el problema primal (1), verifica que* Ω *es convexo y las funciones* $\mathbf{g}, \mathbf{h}$ *son funciones lineales, (*$\mathbf{g}(\mathbf{w}) = \mathbf{A}\mathbf{w} - \mathbf{b}, \mathbf{h}(\mathbf{w}) = \mathbf{B}\mathbf{w} - \mathbf{c}$*), entonces, el gap de dualidad es cero.*

Teorema 16 ***Condiciones de óptimo de Kuhn-Tucker****. Dado un problema primal (1) con dominio* Ω *convexo,* f *diferenciable y* $\mathbf{g}, \mathbf{h}$ *lineales como en el teorema anterior, entonces,* $\mathbf{w}^*$ *es óptimo para el problema si y sólo si existen* $\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$*, tales que:*

$$\begin{aligned}
\nabla_{\mathbf{w}} f(\mathbf{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) &= \mathbf{0} \\
\nabla_{\boldsymbol{\beta}} f(\mathbf{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) &= \mathbf{0} \\
\alpha_i^* g_i(\mathbf{w}^*) &= 0, i = 1 : r \\
g_i(\mathbf{w}^*) &\leq 0, i = 1 : r \\
\alpha_i^* &\geq 0, i = 1 : r
\end{aligned}$$

Las relaciones $\alpha_i^* g_i(\mathbf{w}^*) = 0, i = 1 : r$ se denominan **relaciones de complementariedad de Kuhn-Tucker**, e implican que, para restricciones activas (saturadas), se verifica $\alpha_i^* \geq 0$, mientras que, para restricciones inactivas (no saturadas), $\alpha_i^* = 0$. Asimismo, al igual que con los multiplicadores de Lagrange, α_i^* refleja la tasa de cambio en la función objetivo al variar el término independiente de la restricción $g_i(\mathbf{w}^*)$.

Aplicación a un Programa Cuadrático

Sea el problema (primal):

$$\begin{cases} \min\limits_{\mathbf{w}} \frac{1}{2}\mathbf{w}^T\mathbf{A}\mathbf{w} - \mathbf{b}^T\mathbf{w} \\ \text{sujeto a: } \mathbf{C}\mathbf{w} \leq \mathbf{d} \end{cases}$$

con $\mathbf{A}_{p\times p}, \mathbf{b}_{p\times 1}, \mathbf{C}_{m\times p}, \mathbf{d}_{m\times 1}$. Suponiendo que la región factible no es vacía, planteamos el problema dual:

$$\max_{\boldsymbol{\alpha}\geq 0}\{\min_{w}\{\tfrac{1}{2}\mathbf{w}^T\mathbf{A}\mathbf{w} - \mathbf{b}^T\mathbf{w} + \boldsymbol{\alpha}^T(\mathbf{C}\mathbf{w} - \mathbf{d})\}\}$$

donde $\boldsymbol{\alpha} \in \mathbb{R}^m$ y $\boldsymbol{\alpha} \geq 0$ significa $\alpha_i \geq 0\ \forall i \in \{1 : m\}$.

El mínimo se obtiene para $\mathbf{w} = \mathbf{A}^{-1}(\mathbf{b} - \mathbf{C}^T\boldsymbol{\alpha})$ y sustituyendo esta expresión en el primal, se tiene el dual:

$$\begin{cases} \max\limits_{\boldsymbol{\alpha}} -\tfrac{1}{2}\boldsymbol{\alpha}^T\mathbf{P}\boldsymbol{\alpha} - \boldsymbol{\alpha}^T\mathbf{u} - \tfrac{1}{2}\mathbf{b}^T\mathbf{Q}\mathbf{b} \\ \text{sujeto a: } \boldsymbol{\alpha} \geq 0 \end{cases}$$

donde $\mathbf{P} = \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^T$, $\mathbf{u} = \mathbf{d} - \mathbf{C}\mathbf{A}^{-1}\mathbf{b}$. De esta forma, el dual de un programa cuadrático es otro programa cuadrático pero con restricciones más sencillas.

Bibliografía

Chong E K P y Zak S H (2001). *An Introduction to Optimization*, John Wiley.

Luenberger D E (1984). *Programación Lineal Y No Lineal*, Addison-Wesley Iberoamericana.