

Το δυϊκό πρόβλημα, οι πολλαπλασιαστές Lagrange και ερμηνείες τους

Νίκος Μελανίτης-Παρασκευάς

Σ.Η.Μ.Μ.Υ.

Ιανουάριος, 2013

Αρχικό πρόβλημα βελτιστοποίησης με περιορισμούς:

Πρωτεύον

$$\min f(x) \text{ s.t. } x \in X, g_j \leq 0, j = 1 \dots r$$

Για να λύσουμε το πρόβλημα, χρησιμοποιούμε την Λαγκραντζιανή
: $L(x) = f(x) + \sum_{j=1}^r \mu_j \times g_j, x \in X, \mu_j \geq 0$

Χωρίς υποθέσεις για τις συναρτήσεις f, g και το σύνολο X , μπορούμε να συνδέσουμε τα σημεία σέλας (saddlepoints) της Λαγκρατζιανής και τις λύσεις του αρχικού προβλήματος με σχέσεις ικανές, όχι όμως αναγκαίες. (Karush-Tucker Saddlepoint Problem $\Rightarrow$ Minimisation Problem). Για να καταλήξουμε σε σχέσεις ισοδυναμίας, χρειάζονται υποθέσεις όπως κυρτότητα, παραγωγισιμότητα, constraint qualifications κ.λ.π. (Regular points)

Θα λέμε ότι για κάποιο πρόβλημα, οι πολλαπλασιαστές Lagrange υπάρχουν, αν $\exists \mu^* : f^* = \inf L(x, \mu^*)$

Δυϊκή συνάρτηση

$$q(\mu) = \inf_{x \in X} L(x, \mu)$$

Δυϊκό πρόβλημα

$$\max q(\mu), \mu \geq 0$$

Δύο σημαντικές ιδιότητες του δυϊκού προβλήματος:

- Το σύνολο D στο οποίο η q είναι πεπερασμένη είναι κυρτό
- Η q είναι κοίλη στο σύνολο D

Η βελτιστοποίηση κυρτών ή κοίλων συναρτήσεων σε κυρτά σύνολα είναι εκτενώς μελετημένο πρόβλημα, σημαντικά πιο εύκολο από την βελτιστοποίηση γενικών συναρτήσεων (δεν έχουμε εισάγει περιορισμούς για την f έως τώρα)

Απόδειξη

Έστω $x, \mu, \bar{\mu}$ και $\alpha \in [0, 1]$ έχουμε:

$$L(x, \alpha\mu + (1 - \alpha)\bar{\mu}) = \alpha L(x, \mu) + (1 - \alpha)L(x, \bar{\mu})$$

$$\Rightarrow \inf_{x \in X} L(x, \alpha\mu + (1 - \alpha)\bar{\mu}) \geq$$

$$\inf_{x \in X} \alpha L(x, \mu) + \inf_{x \in X} (1 - \alpha)L(x, \bar{\mu})$$

$$q(\alpha\mu + (1 - \alpha)\bar{\mu}) \geq \alpha q(x, \mu) + (1 - \alpha)q(x, \bar{\mu})$$

Έτσι αν $\mu, \bar{\mu}$ ανήκουν στο D , θα ανήκει και οποιοσδήποτε κυρτός συνδυασμός τους $\alpha\mu + (1 - \alpha)\bar{\mu}$ και η q είναι κοίλη στο σύνολο D .



Ασθενής δυϊκότητα

$$q^* \leq f^*$$

Η διαφορά $f^* - q^*$ ονομάζεται duality gap. Για περιπτώσεις στις οποίες το σημείο του ελαχίστου της f είναι ομαλό (regular point), δηλαδή υπάρχουν πολλαπλασιαστές Lagrange στο σημείο ελαχίστου της f , το κενό δυϊκότητας είναι 0. Όμως η ύπαρξη πολλαπλασιαστών Lagrange δεν είναι αναγκαία συνθήκη.

Απόδειξη

$$\forall \mu \geq 0, \mathbf{x} \in X$$

$$q(\mu) = \inf_{z \in X} L(z, \mu) \leq f(\mathbf{x}) + \sum_{j=1}^r \mu_j g_j(\mathbf{x}) \leq f(\mathbf{x})$$

Η πρώτη δεύτερη ανισότητα ακολουθεί από την σχέση

$$g(\mathbf{x}) \leq 0, \bar{\mathbf{x}} \in X$$

$$q^* = \sup_{\mu \geq 0} q(\mu) \leq \inf_{\mathbf{x} \in X, g(\mathbf{x}) \leq 0} f(\mathbf{x}) = f^*$$



Ένα μικρό παράδειγμα για το duality gap

$$\min x, \text{ s.t. } g(x) = x^2 \leq 0$$

$$L(x, \mu) = x + \mu x^2$$

$$\inf_x L = \inf_x (x + \mu x^2) = q(\mu)$$

$$q^* = \sup_{\mu \geq 0} q(\mu) = 0$$

με μαθηματικά λυκείου για $\mu > 0$ (Μελέτη τριωνύμου)

για $\mu = 0$, $q(0) = -\infty$

Το supremum επιτυγχάνεται για $\mu \rightarrow \infty$. Δεν υπάρχει επομένως πολλαπλασιαστής Lagrange (επίπεδο κάθετο). Duality gap = 0.

Πρόταση

- Αν το κενό δυϊκότητας είναι 0, το σύνολο των πολλαπλασιαστών Lagrange ταυτίζεται με το σύνολο βέλτιστων δυϊκών λύσεων
- Αν υπάρχει κενό δυϊκότητας, το σύνολο των πολλαπλασιαστών Lagrange είναι κενό

Από την προηγούμενη πρόταση απορρέει η μη ύπαρξη βέλτιστης δυϊκής λύσης στις περιπτώσεις μηδενικού κενού δυϊκότητας χωρίς την ύπαρξη πολλαπλασιαστών Lagrange.

Περιπτώσεις με ισχυρή δυϊκότητα:

- Σύνολο X πολύεδρο και οι περιορισμοί είναι γραμμικοί
- Οι περιορισμοί είναι κυρτοί και $\exists x : g_j(x) < 0, \forall j$, με x στο εσωτερικό του X . (Slater constraint qualification)

Παρατηρούμε ότι $\sup_{\mu \geq 0} L(x, \mu) = f(x)$ αν το x είναι εφικτό (feasible) σημείο του προβλήματος και αλλιώς είναι ∞ . Μπορούμε επομένως να διατυπώσουμε το πρωτεύον πρόβλημα σε συμμετρική μορφή με το δυϊκό και έχουμε:

- Ασθενής δυϊκότητα:

$$p^* \geq d^* \iff \inf_x \sup_{\mu \geq 0} L(x, \mu) \geq \sup_{\mu \geq 0} \inf_x L(x, \mu)$$

(max-min inequality)

- Ισχυρή δυϊκότητα:

$$p^* = d^* \iff \inf_x \sup_{\mu \geq 0} L(x, \mu) = \sup_{\mu \geq 0} \inf_x L(x, \mu)$$

(saddlepoint property)

Εστώ x^*, μ^* η βέλτιστη λύση στο πρωτεύον και δυϊκό πρόβλημα αντίστοιχα.

- Αν ισχύει η ισχυρή δυϊκότητα, το σημείο (x^*, μ^*) είναι σημείο σέλας της Λαγκραντζιανής
- Αν το (x, μ) είναι σημείο σέλας της Λαγκραντζιανής, τότε x :primal optimal, μ :dual optimal και το κενό δυϊκότητας είναι 0.

Απόδειξη

Αν (x^*, μ^*) είναι αντίστοιχα βέλτιστες λύσεις, τότε $x^* = \operatorname{argmin}_{x \in X} L(x, \mu^*)$, οπότε $L(x^*, \mu^*) \leq L(x, \mu^*)$. Επίσης επειδή $g(x^*) \leq 0$ και $\mu \geq 0$ έχουμε $L(x^*, \mu) \leq f(x^*) = L(x^*, \mu^*)$ και το πρώτο μέρος της απόδειξης ολοκληρώθηκε.

Για το δεύτερο σκέλος, έχουμε

$$L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*), \forall x \in X, \mu \geq 0.$$

Προκειμένου $\sup_{\mu \geq 0} L(x^*, \mu) < \infty$, ώστε να ισχύει το αριστερό μέλος της ανισότητας, πρέπει $g(x^*) \leq 0$, όπως αναλύθηκε προηγούμενα. Έτσι $f(x^*) = L(x^*, \mu^*)$ και

$\sum_{j=1}^r \mu_j^* g_j(x^*) = 0 \Rightarrow \mu_j^* g_j(x^*) = 0$. Από το δεξί μέλος της ανισότητας $x^* = \operatorname{argmin}_{x \in X} L(x, \mu^*)$ και η αποδείξή ολοκληρώθηκε



Λήμμα

Ορίζουμε το σύνολο $S = \{(g(x), f(x)) \mid x \in X\}$. Ένα υπερεπίπεδο με κάθετο διάνυσμα $(\mu, 1)$ ορίζεται από την εξίσωση $(\mu, 1) * (w, z)' = c$.

- Το υπερεπίπεδο με κάθετο διάνυσμα $(\mu, 1)$ το οποίο διέρχεται από το σημείο $(g(x), f(x))$ τέμνει τον κάθετο άξονα $(0, w)$ στη στάθμη $L(x, \mu)$
- Ανάμεσα σε όλα τα υπερεπίπεδα με κάθετο $(\mu, 1)$ που έχουν στον θετικό τους ημιχώρο το σύνολο S , το μέγιστο σημείο τομής με τον κάθετο άξονα είναι $\inf_{x \in X} L(x, \mu)$

Απόδειξη

Για το δεύτερο σκέλος έχουμε:

Εξίσωση υπερεπιπέδου: $w + \mu'z = c$, τέμνει τον κάθετο άξονα στο επίπεδο $z = c$

Για να ανήκει στο θετικό ημιχώρο του επιπέδου το σύνολο S πρέπει: $L(x, \mu) = f(x) + \mu' * g(x) \geq c, \forall x \in X$

Επομένως $\max c = \inf_{x \in X} L(x, \mu)$

Κάθε υπερεπίπεδο που αφήνει το σύνολο S στον θετικό του υποχώρο, ορίζει ένα κάτω φράγμα για την βέλτιστη λύση. Μεταβάλλοντας την τιμή του μ αλλάζουμε την κλίση του υπερεπιπέδου. Λύνοντας το δυϊκό πρόβλημα βρίσκουμε το καλύτερο κάτω φράγμα, δηλαδή το μέγιστο σημείο τομής υπερεπιπέδου με τον κάθετο άξονα. Η μορφή του συνόλου S καθορίζει την ύπαρξη κενού δυϊκότητας.

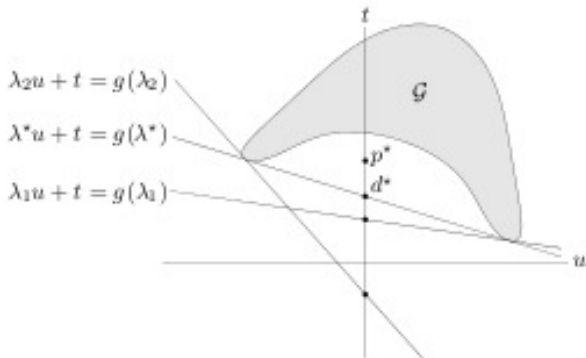
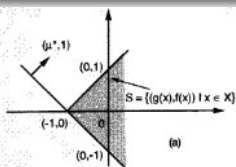
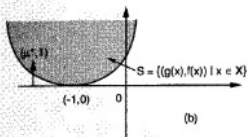


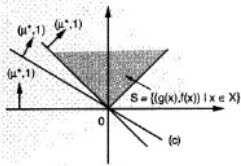
Figure 5.4 Supporting hyperplanes corresponding to three dual feasible values of λ , including the optimum λ^* . Strong duality does not hold; the optimal duality gap $p^* - d^*$ is positive.



$$\begin{aligned} \min f(x) &= x_1 \cdot x_2 \\ \text{s.t. } g(x) &= x_1 + x_2 - 1 \leq 0 \\ x \in X &= \{(x_1, x_2) \mid x_1 \geq 0, x_2 \geq 0\} \end{aligned}$$

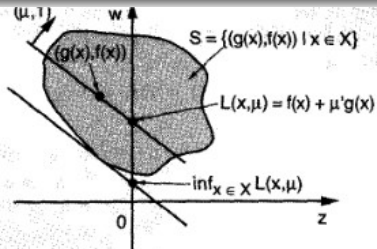


$$\begin{aligned} \min f(x) &= (1/2)(x_1^2 + x_2^2) \\ \text{s.t. } g(x) &= x_1^2 + x_2^2 - 1 \leq 0 \\ x \in X &= \mathbb{R}^2 \end{aligned}$$

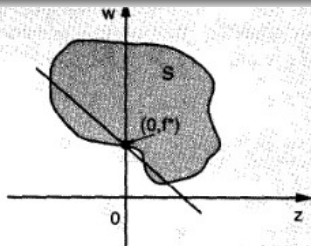


$$\begin{aligned} \min f(x) &= |x_1| + x_2 \\ \text{s.t. } g(x) &= x_1 \leq 0 \\ x \in X &= \{(x_1, x_2) \mid x_2 \geq 0\} \end{aligned}$$

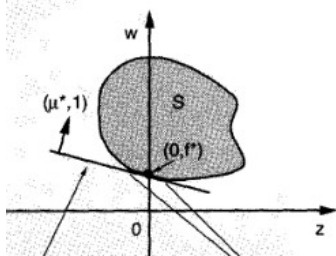
Figure 5.1.3. Examples where there exists at least one Lagrange multiplier. In (a), there is a unique Lagrange multiplier, $\mu^* = 1$. In (b), there is a unique Lagrange multiplier, $\mu^* = 0$. In (c), the set of Lagrange multipliers is the interval $[0, 1]$.



(a)



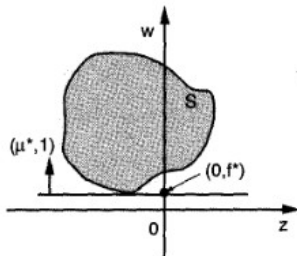
(b)



$$H = \{(z, w) \mid f^* = w + \sum_j \mu_j^* z_j\}$$

Set of pairs $(g(\bar{x}), f(\bar{x}))$ corresponding to $\bar{x}$ that minimize $L(x, \mu^*)$ over X

(c)



(d)

Παρατηρούμε ότι στην περίπτωση παράλληλου επιπέδου, δηλ $\mu = 0$, η ανισότητα δεν ικανοποιείται με ισότητα στο σημείο ελαχίστου. Η σχέση αυτή μας θυμίζει την ιδιότητα complementary slackness, η οποία μπορεί πράγματι ναδειχθεί. Όμως πρέπει να ισχύει η ισχυρή δυϊκότητα (duality gap 0).

Στα επόμενα υποθέτουμε διαφορίσιμες συναρτήσεις, καθώς και ότι η επόμενη πρόταση ισχύει, η οποία εξασφαλίζει την ύπαρξη πολλαπλασιαστών Lagrange στο σημείο ελαχίστου της f .

Ικάνες συνθήκες δεύτερης τάξης

Υποθέτουμε f, h, g διπλά συνεχώς διαφορίσιμες και ότι μ^*, λ^*, x^* ικανοποιούν :

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, h(x^*) = 0, g(x^*) \leq 0$$

$$\mu_j^* \geq 0, j = 1 \cdots r$$

$$\mu_j^* = 0, \forall j \notin A(x^*)$$

$$y' \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) y > 0$$

για κάθε $y \neq 0$ τέτοιο ώστε:

$$\nabla h_i(x^*)' y = 0, \forall i, \nabla g_j(x^*)' y = 0, \forall j \in A(x^*)$$

Τότε το x^* είναι τοπικό ελάχιστο.

Το παρακάτω θεώρημα μας επιτρέπει να ερμηνεύσουμε τους πολλαπλασιαστές Lagrange ως "τιμές" για την χαλάρωση-παραβίαση των περιορισμών. Έτσι μπορούμε να μελετήσουμε οφέλη από την διατύπωση προβλήματος ελαφρά διαφορετικού (perturbed) από το δοθέν.

Θεώρημα Ευαισθησίας

Θεώρουμε x^* , λ^* , σημείο τοπικού ελαχίστου και αντίστοιχοι πολλαπλασιαστές Lagrange οι οποίοι ικανοποιούν την προηγούμενη ικανή δευτεροβάθμια συνθήκη. Θεωρούμε την οικογένεια προβλημάτων :

$$\min f(x) \text{ s.t. } h(x) = u, g(x) \leq v$$

Τότε ισχύουν: $\nabla_u p(u, v) = -\lambda(u, v)$, $\nabla_v p(u, v) = -\mu(u, v)$
όπου $p(u, v) = f(x(u, v))$ είναι το βέλτιστο κόστος
περαμετρικοποιημένο από τα u, v

Παράδειγμα αντί για απόδειξη

Έστω $f = f(x_1, x_2)$, $g(x_1, x_2) = c$. Στα κρίσιμα σημεία έχουμε $\nabla f = \lambda \nabla g \Rightarrow \frac{\partial f}{\partial x_1} = \lambda \frac{\partial g}{\partial x_1}$, $\frac{\partial f}{\partial x_2} = \lambda \frac{\partial g}{\partial x_2}$, $g = c$.

Υποθέτουμε ότι έχουμε ομαλή λύση (regular point), όποτε το ανωτέρω σύστημα μας δίνει λύση έστω $\hat{x}_1, \hat{x}_2, \hat{\lambda}$ και βέλτιστο κόστος $\hat{f}$. Παραγωγίζουμε την f ως προς το επίπεδο του περιορισμού c και έχουμε διαδοχικά, χρησιμοποιώντας και την προηγούμενη σχέση:

$$\frac{d\hat{f}}{dc} = \frac{df(\hat{x}_1(c), \hat{x}_2(c))}{dc} = \frac{\partial f}{\partial x_1} \frac{d\hat{x}_1}{dc} + \frac{\partial f}{\partial x_2} \frac{d\hat{x}_2}{dc} = \lambda \frac{\partial g}{\partial x_1} \frac{d\hat{x}_1}{dc} + \lambda \frac{\partial g}{\partial x_2} \frac{d\hat{x}_2}{dc} = \lambda \left(\frac{\partial g}{\partial x_1} \frac{d\hat{x}_1}{dc} + \frac{\partial g}{\partial x_2} \frac{d\hat{x}_2}{dc} \right) = \lambda \times 1$$

Η ποσότητα εντός της παρένθεσης ισούται με 1 γιατί

$$1 = \frac{d\hat{g}}{dc} = \frac{\partial g}{\partial x_1} \frac{d\hat{x}_1}{dc} + \frac{\partial g}{\partial x_2} \frac{d\hat{x}_2}{dc}$$

Ένα Αριθμητικό Παράδειγμα

Θεωρούμε το εξής πρόβλημα. Έστω ότι διαθέτουμε συγκεκριμένη ποσότητα του αγαθού "φράχτης". Θέλουμε να τον τοποθετήσουμε ώστε να μεγιστοποιήσουμε το εμβαδόν του περιφραγμένου οικοπέδου, ενώ θεωρούμε πως η περίφραξη μας είναι ορθογώνιο. Γράφουμε τις εξισώσεις του προβλήματος και έχουμε:

$$\max f(x, y) = xy, \text{ s.t. } 2x + 2y \leq F.$$

Από την σχέση $\nabla L(\underline{x}, \lambda) = \vec{0}$ βρίσκουμε $x = y = \frac{F}{4}, \lambda = \frac{F}{8}$

- Συμπέρασμα: Αν ο αγρότης θέλει να αυξήσει τον φράχτη κατά 1, η περιφραγμένη επιφάνεια θα αυξηθεί κατά $\frac{F}{8}$
- Για περίμετρο $F = 400 \Rightarrow \text{Εμβαδόν} = 10.000$. Προσεγγιστικά, αν $F_{\nu} = 401 \Rightarrow S_{\nu} \approx 10.050$. (Ακριβής υπολογισμός δίνει $\frac{401^2}{4} = 10.050,06$)
- Γενικά μπορούμε να υπολογίσουμε αν η επένδυση που χαλαρώνει κάποιον περιορισμό θα προσκομίσει ικανοποιητικά οφέλη (μεταβολές στην συνάρτηση κόστους)
- Shadow prices: Αν οι περιορισμοί εκφράζουν ποσότητες αγαθών που μπορούμε να αγοράσουμε/πουλήσουμε τότε τα λ^* είναι οι τιμές ισοροπίας τους. Αν μπορούσαμε να αγοράσουμε ένα αγαθό σε τιμή μικρότερη από λ^* τότε η βελτίωση της συνάρτησης κόστους θα ήταν μεγαλύτερη (δηλαδή συμφέρουσα) από το κόστος αγοράς του αγαθού.

Standard form L.P.:

$$\begin{array}{ll}\min & c'x \\ \text{s.t.} & Ax \leq b\end{array}$$

Η Λαγκραντζιανή του προβλήματος είναι:

$$L(x, \lambda) = c'x + \lambda'(Ax - b) = -b'\lambda + (A'\lambda + c)'x$$

Η δυϊκή συνάρτηση κόστους είναι

$$g(\lambda) = \inf_x L(x, \lambda) = -b'\lambda + \inf_x (A'\lambda + c)'x$$

δηλαδή γραμμική συνάρτηση του x , οπότε είναι φραγμενη όταν ο συντελεστής του x μηδενίζεται.

$$g(\lambda) = \begin{cases} -b'\lambda, & A'\lambda + c = 0 \\ -\infty, & \text{αλλιώς} \end{cases}$$

Αν ενσωματώσουμε στο δυϊκό πρόβλημα τον περιορισμό που εξασφαλίζει ότι η δυϊκή συνάρτηση δεν αποκλίνει, καταλήγουμε στην γνωστή μορφή:

$$\max -b'\lambda$$

$$\text{s.t. } A'\lambda + c = 0, \lambda \geq 0$$

-  D. Bertsekas, Nonlinear Programming, Athena Scientific, Massachusetts, 2nd Edition, 1999
-  S. Boyd L. Vandenberghe, Convex Optimisation, Cambridge University Press, 2004
-  O. Mangasarian, Nonlinear Programming, S.I.A.M., 1987
-  I. Meznik, On economic interpretation of Lagrange multipliers, 2011