Μηχανική Μάθηση

Ενότητα 10: Θεωρία Βελτιστοποίησης

Ιωάννης Τσαμαρδίνος
Τμήμα Επιστήμης Υπολογιστών

Το γενικό πρόβλημα, να βρούμε το μέγιστο ή ελάχιστο μιας συνάρτησης $f(x)$

Γενικα,

$$\begin{array}{lll} minimize & f(w), & w \in \Omega \subseteq R^n \\ s.t & g_i(w) \leq 0 & i = 1, \ldots, k \\ & h_i(w) = 0 & i = 1, \ldots, m \end{array}$$

f(w)	objective function
$g_i(w)$	inequality constraints
$h_i(w)$	equality constraints

Maximization of f(w) $\rightarrow$ minimization of -f(w)

Κάθε πρόβλημα μπορεί να γραφεί στην παραπάνω μορφή

- Το σύνολο $R = \{w \in \Omega : g_i(w) \leq 0, h_i(w) = 0\}$ ονομάζεται feasible region
- Η λύση $f(w^\star)$ με $w^\star \in R$, έτσι ώστε $f(w) \geq f(w^\star), \quad w \in R$ ονομάζεται καθολικό ελάχιστο και το $w^\star$ ο ελαχιστοποιητής
- Αν $\exists \varepsilon > 0, f(w) \geq f(w^\star), w \in R$ με $\|w - w^\star\| < \varepsilon$ τότε $f(w^\star)$ τοπικό ελάχιστο

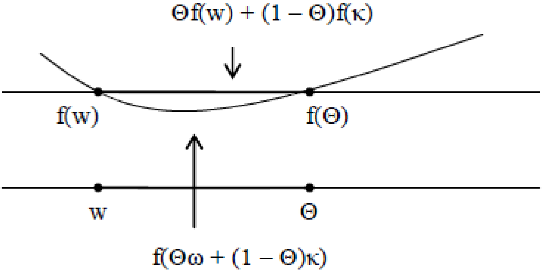
Το πεδίο ονομάζεται μαθηματικός προγραμματισμός.

- Αν f, g, h γραμμικές
    - Γραμμικός Προγραμματισμός
    - Λύση πολυωνυμική
- Αν f, quadratic, g, h γραμμικές
    - Quadratic Programming
    - $f(x) = x^T Qx + c^T x$
    - Αν X θετικά ημιορισμένος (positive semidefinite), μοναδική λύση, πολυωνυμική λύση
    - Αν X αρνητικά ημιορισμένος (negative semidefinite), NP-hard

(Q θετικά ημιορισμένος αν είναι συμμετρικός και για κάθε $x \neq 0$ ισχύει ότι $x^T Qx \geq 0$)

- Μια συνεχής συνάρτηση $f(w)$ ονομάζεται κυρτή για $w \in R^n$ αν :

$$\forall w, \kappa \in R^n, \forall \Theta \in [0,1] : \quad f(\Theta w + (1-\Theta)\kappa) \leq \Theta f(w) + (1-\Theta) f(\kappa)$$



Ένα σύνολο είναι κυρτό (convex) αν το ευθύγραμμο τμήμα μεταξύ δυο σημείων του ανήκει στο σύνολο:

$$R_{convex} \Rightarrow \forall w, \kappa \in R, \forall \Theta \in [0,1] \quad \Theta f(w) + (1\text{–}\Theta) f(\kappa) \in R$$

- Ιδιότητες:
  1. Αν S κυρτό, $\kappa_1, \ldots, \kappa_n \in S, \lambda_1, \ldots, \lambda_n \leq 0, \lambda_1 + \cdots + \lambda_n = 1$ τότε $\sum_i \lambda_i \kappa_i$: κυρτός συνδυασμός των κ_i
  2. Τομή κυρτών συνόλων είναι κυρτή.
  3. Μια συνάρτηση f είναι κυρτή αν και μόνο αν το επίγραμμά της είναι κυρτό.

Αν f κυρτή και R κυρτό $\rightarrow$ convex programming optimization

- Ιδιότητες:
  1. Αν υπάρχει ελάχιστο είναι καθολικό.
  2. Το σύνολο των ελαχίστων είναι convex.
  3. Αν είναι strictly convex $\rightarrow$ μοναδικό ελάχιστο.

Πρακτικά, πολλά πακέτα και αλγόριθμοι σχετικά γρήγοροι.

Θεώρημα (Fermat):
Έστω $f \in C^\star$ (έχει παράγωγο και η παράγωγος είναι συνεχής).
Απαραίτητη συνθήκη για το $w^\star$ να είναι ο ελαχιστοποιητής της $f(w^\star)$ είναι:

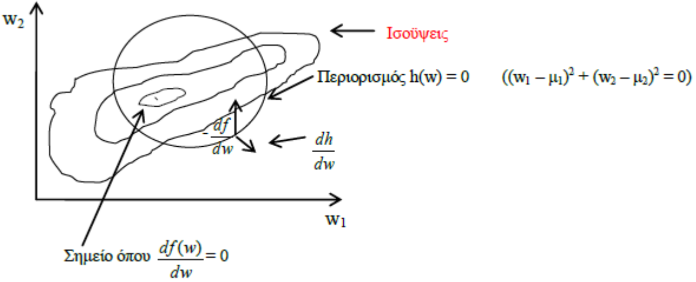
$$\frac{\partial f(w^\star)}{\partial w} = 0$$

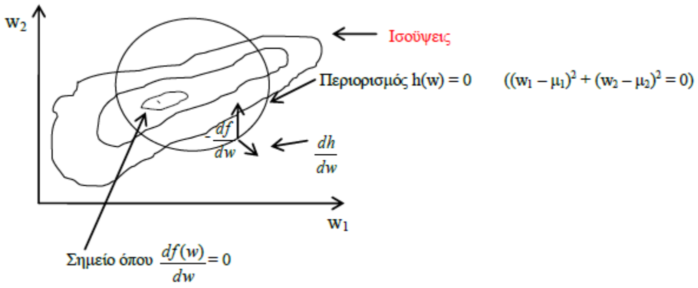
Αν η f είναι convex, η συνθήκη είναι και ικανή π.χ.

- $f(x) = x^2$
- $f(x) = x^3$

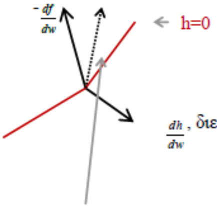
Με περιορισμούς:

- $min\ f(w)$
- $s.t.\ h_i(w) = 0$



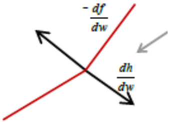


Για να μείνουμε στην περιοχή όπου $h(w) = 0$, πρέπει να κινηθούμε παράλληλα με αυτήν, δηλαδή κάθετα στο $\frac{\partial h}{\partial w}$ Γενικά πρέπει να κινηθούμε κάθετα στον υποχώρο $\{\frac{\partial h_i}{\partial w}\}$
Αν μπορούμε να κινηθούμε σε αυτόν τον υπόχωρο μειώνοντας ταυτόχρονα την τιμή f τότε δεν είμαστε σε ελάχιστο. Αν όμως το $\frac{\partial f}{\partial w}$ δεν έχει προβολή πάνω σε αυτόν τον χώρο, θα πρέπει να είμαστε σε κάποιο τοπικό ελάχιστο.



$\frac{dh}{dw}$, διεύθυνση που η h αυξάνει το περισσότερο

σε αυτή τη διεύθυνση μειώνουμε την f και παραμένει h = 0



Δε γίνεται να μειώσουμε κι άλλο την f
και να μείνουμε στο σύνολο h = 0

Άρα θέλουμε $\frac{\partial f}{\partial w}$ παράλληλο του $\frac{\partial h}{\partial w}$, στο ελάχιστο.
Δηλαδή:

$$\frac{\partial f(w^\star)}{\partial w} = \lambda \frac{\partial h(w^\star)}{\partial w}$$

Δηλαδή θέλουμε:

$$\frac{\partial (f(w^\star) + \lambda h(w^\star))}{\partial w} = 0$$

και φυσικά:

$$h(w^\star) = 0$$

δηλαδή:

$$\frac{\partial \lambda h(w^\star)}{\partial \lambda w} = 0 \qquad \text{ή} \qquad \frac{\partial (w^\star) + \lambda h(w^\star}{\partial \lambda} = 0$$

Συνολικά μπορούμε να τα εκφράσουμε με τη συνάρτηση Lagrangian:

$$L(w,\lambda) = f(w) + \textstyle\sum_i \lambda_i h_i(w)$$
$$\lambda_i : \text{Lagrange multipliers}$$

Θεώρημα: Έστω $f \in C^1, \quad h_i \in C^1$ Αναγκαία συνθήκη ώστε το $w^\star$ να είναι ελάχιστο είναι:

$$\left.\begin{array}{l} \frac{\partial L(w^\star,\lambda^\star)}{\partial w^\star} = 0 \\ \\ \frac{\partial L(w^\star,\lambda^\star)}{\partial \lambda^\star} = 0 \end{array}\right. \Leftrightarrow \nabla L = 0 \text{ για κάποιες τιμές} \lambda^\star.$$

(σαν να ελαχιστοποιούμε την L χωρίς περιορισμούς)
Αν $L(w,\lambda)$ convex, η συνθήκη είναι και ικανή.
Η (1) δίνει ένα σύστημα εξισώσεων (όση η διάσταση του w)
Η (2) (όσα τα constraints).
Από μια συνάρτηση n διαστάσεων με περιορισμούς, ελαχιστοποιούμε μια με $n+m$ διαστάσεις χωρίς περιορισμούς

Παράδειγμα: Ποια κατανομή με n τιμές έχει τη μέγιστη εντροπία;

$$\begin{array}{ll} min & -H(x) = \sum_{x_i} p_i log(p_i) \\ s.t. & \quad \sum_i p_i = 1, p_i \geq 0 \end{array}$$

Λύση:

$$L(p,\lambda) = \sum p_i log(p_i) + \lambda(\sum p_i - 1)$$

$$\frac{\partial L(p,\lambda)}{\partial p_j} = p_j \frac{1}{p_j} + log(p_j) + \lambda = 0 \Rightarrow p_j = e^{-\lambda - 1}$$

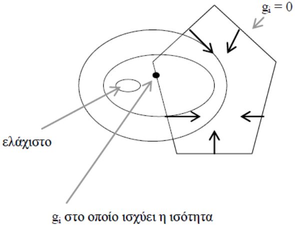
$$\frac{\partial L(p,\lambda)}{\partial \lambda} = \sum p_i - 1 = 0 \Rightarrow n \cdot e^{-\lambda - 1} - 1 = 0 \Rightarrow e^{-\lambda - 1} - 1 = \frac{1}{n} \Rightarrow p_j = \frac{1}{n}$$

Πιο γενική μορφή:

$$\begin{aligned} &\textit{minimize} && f(w) \\ &s.t. && g_i(w) \leq 0 \\ &&& h_i(w) = 0 \end{aligned}$$

Επίσης f convex
g_i, h_i affine
(γραμμικές)

Για τον ελαχιστοποιητή $w^\star$
$g_i(w^\star) = 0$: ο περιορισμός g_i είναι ενεργός
$g_i(w^\star) > 0$: ο περιορισμός g_i είναι ανενεργός
Άρα οι ενεργοί περιορισμοί ανισότητας στην λύση $w^\star$ καταληγουν να ειναι περιορισμοί ισότητας και μπορούμε να τους χειριστούμε αντίστοιχα. Αλλά δεν ξέρουμε ποιοί ειναί οι ενεργοί περιορισμοί.



Karush – Kuhn – Tucker
Θεώρημα: $f \in C, w \in \Omega, f$ convex , g_i, h_i affine.
Ορίζουμε ως Γενικευμένη Lagrangian:

$$L(w, \alpha, \beta) = f(w) + \sum \alpha_i g_i(w) + \sum \beta_i h_i(w)$$

Αναγκαία και ικανές συνθήκες ώστε $w^\star \in \Omega$ ελαχιστοποιητής είναι: $\exists \alpha^\star, \beta^\star$ s.t. :

1. $\frac{\partial L(w^\star, \alpha^\star, \beta^\star)}{\partial w} = 0$
2. $\frac{\partial L(w^\star, \alpha^\star, \beta^\star)}{\partial \beta} = 0$
3. $\alpha_i^\star \cdot g_i(w^\star) = 0$
4. $g_i(w^\star) \leq 0$
5. $\alpha_i^\star \geq 0$

$(2) + (4) : \Leftrightarrow \; w^\star$ feasible point
$(1) + (3) + (5)$: $\Leftrightarrow$ Παρόμοιοι περιορισμοι για τα active constraints με αυτούς που είχαμε για τα h.

1. Αν $\alpha_i^\star > 0 \Rightarrow g_i(w^\star) = 0$ (από την (3)), g_i active constraint
2. Αν $g_i(w^\star) < 0$ (Μη ενεργά constraints) $\Rightarrow \alpha_i^\star = 0$
3. Ενδέχεται $g_i(w^\star) = 0$ και $\alpha_i^\star = 0$

Έστω η συνάρτηση

$$\gamma(\alpha,\beta) = inf_{w\in\Omega} L(w,\alpha,\beta) =$$

$$= f(w) + \sum \alpha_i g_i(w) + \sum \beta_i h_i(w)$$

με $\alpha_i \geq 0$, f convex, g_i, h_i affine.

Langrange dual

$max \quad \gamma(\alpha,\beta)$ s.t. $\alpha_i \geq 0$

α, β feasible w feasible, τότε

$$\gamma(\alpha,\beta) = min_{u\in\Omega} L(u,\alpha,\beta) \le L(w,\alpha,\beta) =$$

$$= \left. \begin{array}{lll} f(w)+ & \sum \alpha_i g_i(w)+ & \sum \beta_i h_i(w) \\ & \text{w feasible} & \Rightarrow g_i(w) \le 0 \\ & & \Rightarrow h_i(w) = 0 \end{array} \right\} \longrightarrow \le f(w)$$

Άρα $f(w) \ge \gamma(\alpha,\beta)$
Δηλαδή, για κάθε feasible $\gamma(\alpha,\beta)$ lower bound σε κάθε feasible $f(w)$

$$\text{Άρα} \quad \begin{array}{ll} max & \gamma(\alpha,\beta) \\ s.t & \alpha \ge 0 \end{array} \quad \le \quad \begin{array}{ll} min & f(w) \\ & g(w) \le 0 \\ & h(w) = 0 \end{array}$$

Για τις συνθήκες που δώσαμε ισχύει η ισότητα (strong duality).

$$\gamma(\alpha^\star, \beta^\star) = f(w^\star), \quad (\alpha^\star, \beta^\star, w^\star \text{λύσεις})$$

- Το dual problem λύνει το min $L(w, \alpha, \beta)$ unconstrained (με $\alpha \geq 0$) βρίσκοντας $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \beta} = 0$
- Λύνοντας ως προς w το $\frac{\partial L}{\partial w} = 0$, μένουν μόνο τα $\alpha, \beta \to$ max $\gamma(\alpha, \beta)$ s.t. $\alpha \geq 0$, και ˙ετσι το πρόβλημα μπορεί να είναι πιο εύκολο.
- Αν το unconstraint min$L(w, \alpha, \beta)$ λύνεται αναλυτικά τότε το dual έχει διάσταση όση το $\alpha + \beta$
- Για κάθε πρόβλημα βελτιστοποίησης, τα primal και dual βέλτιστα σημεία ικανοποιούν τις KKT conditions
- Για κάθε convex πρόβλημα, αν κάποια σημεία ικανοποιούν τα KKT είναι βέλτιστα και για το primal και για το dual.

Θέλω να λύσω το

$$\max_{\alpha,\beta} \quad \gamma(\alpha^\star, \beta^\star) \quad \text{s.t. } \alpha \geq 0$$

Δηλαδή (απο ορισμο γ)

$$\max_{\alpha,\beta} \quad min_w \quad L(w, \alpha, \beta) \quad \text{s.t. } \alpha \geq 0$$

Το $min_w \quad L(w, \alpha, \beta)$ γίνεται εκεί όπου

$$\frac{\partial L}{\partial w} = 0, \frac{\partial L}{\partial \beta} = 0$$

Άρα Ισοδύναμα λύνω το(Wolfe dual)

$$\max_{\alpha} \quad L(w, \alpha, \beta) \quad \text{s.t. } \alpha \geq 0, \frac{\partial L}{\partial w} = 0, \frac{\partial L}{\partial \beta} = 0$$

Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημειώματα

Σημείωμα αδειοδότησης (1)

Σημείωμα αδειοδότησης (2)

- Ως **Μη Εμπορική** ορίζεται η χρήση:
  - που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
  - που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
  - που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο
- Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Ιωάννης Τσαμαρδίνος 2015. «Μηχανική Μάθηση. Θεωρία Βελτιστοποίησης». Έκδοση: 1.0. Ηράκλειο 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: https://opencourses.uoc.gr/courses/course/view.php?id=362.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.