Στοιχεία Κυρτής Βελτιστοποίησης

Αθανάσιος Π. Λιάβας
Τμήμα ΗΜΜΥ
Πολυτεχνείο Κρήτης
Email: liavas@telecom.tuc.gr

Εισαγωγή

Οι παρούσες σημειώσεις αποτελούν υλικό το οποίο αναπτύσεται κατά τη διδασκαλία του μαθήματος "Κυρτή Βελτιστοποίηση." Η Βελτιστοποίηση είναι μία περιοχή των Εφαρμοσμένων Μαθηματικών με εξαιρετικά μεγάλη επίδραση στις Επιστήμες Μηχανικών, στην Οικονομία, στην Επιχειρησιακή Έρευνα κτλ.

Η περιοχή της Κυρτής Βελτιστοποίησης είναι εξαιρετικά ενεργή επιστημονικά, με πληθώρα συνεισφορών από μέλη διαφόρων επιστημονικών κοινοτήτων, όπως Μαθηματικούς, Μηχανικούς, Οικονομολόγους.

Σε αυτό το μάθημα, θα καλύψουμε ένα σχετικά μικρό υποσύνολο της περιοχής, το οποίο θα μπορούσε να χαρακτηριστεί ως εισαγωγικό στην περιοχή της κυρτής βελτιστοποίησης. Η κάλυψη θα περιλαμβάνει

1. συνοπτική επανάληψη εννοιών θεωρίας πραγματικών συναρτήσεων,
2. στοιχεία θεωρίας κυρτών συνόλων και κυρτών συναρτήσεων,
3. ορισμός βασικών εννοιών προβλημάτων (κυρτής) βελτιστοποίησης,
4. εξαγωγή συνθηκών βελτιστότητας,
5. ορισμός προβλημάτων κυρτής βελτιστοποίησης χωρίς περιορισμούς και οι αλγόριθμοι πιο απότομης καθόδου και Newton,
6. ορισμός προβλημάτων κυρτής βελτιστοποίησης με γραμμικούς περιορισμούς και επίλυση μέσω γενίκευσης του αλγορίθμου Newton,
7. ορισμός προβλημάτων κυρτής βελτιστοποίησης και επίλυση με τη μέθοδο εσωτερικού σημείου.

Επίσης, θα καλυφούν μερικές σημαντικές εφαρμογές και, για λόγους πληρότητας, θα παρουσιαστεί η μέθοδος simplex για την επίλυση προβλημάτων γραμμικού προγραμματισμού.

Στη μορφοποίηση του παρόντος υλικού έχουν επιδράσει σημαντικά τα εξής εξαιρετικά βιβλία:

1. S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.
2. M. Bazaraa, H. Sherali, C. Shetty. Nonlinear Programming. Theory and Algorithms. Wiley, 2nd Edition, 1993.

Η σχετική βιβλιογραφία είναι εκτενέστατη και δεν θα αναφερθούμε σε αυτή διεξοδικά. Απλά, θεωρούμε ότι ειδική μνεία πρέπει να γίνει στα εξής:

1. A. Beck. Introduction to Nonlinear Optimization. SIAM, 2014.

2. D. Bertsekas. Nonlinear Programming. Athena Scientific, 2nd Edition, 1999.

Τα σχόλια και οι παρατηρήσεις σας είναι εξαιρετικά ευπρόσδεκτα.

Ο διδάσκων

Αθανάσιος Π. Λιάβας

Κεφάλαιο 2

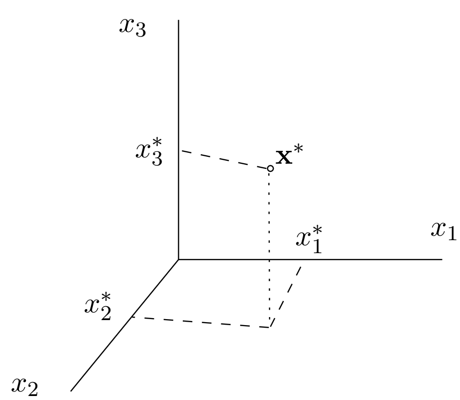
Ευκλείδειοι χώροι

Σε αυτό το Κεφάλαιο θα παρουσιάσουμε, συνοπτικά, βασικά στοιχεία της θεωρίας Ευκλείδειων χώρων. Το βιβλίο [;] περιέχει μία εκτενή περιγραφή και αποτελεί μία εξαιρετική πηγή. Η πολύ καλή γνώση της ύλης η οποία θα παρουσιαστεί συνοπτικά στη συνέχεια θεωρείται απαραίτητη για την κατανόηση του μαθήματος.

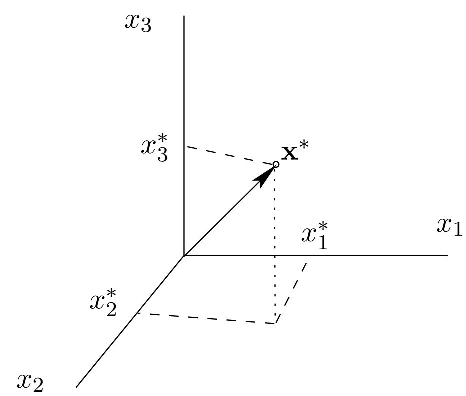
Υποθέτουμε ότι ο αναγνώστης έχει εξοικείωση με τον n-διάστατο Ευκλείδειο χώρο $\mathbb{R}^n$, με $n \in \mathbb{N}$. Αν $n = 1$, τότε ο αντίστοιχος χώρος θα συμβολίζεται ως $\mathbb{R}$. Στη συνέχεια του Κεφαλαίου, γενικά, υποθέτουμε ότι δουλεύουμε στον $\mathbb{R}^n$.

2.1 Σημεία - Διανύσματα

Ορισμός 2.1.1. **Διάνυσμα** (vector) καλείται κάθε διατεταγμένη n-άδα, $(x_1, x_2, \ldots, x_n)$, με $x_i \in \mathbb{R}$, για $i = 1, \ldots, n$.



Σχήμα 2.1: Σημείο στον τριδιάστατο χώρο.



Σχήμα 2.2: Διάνυσμα στον τριδιάστατο χώρο.

Κάθε διάνυσμα θα συμβολίζεται με τρεις ισοδύναμους τρόπους. Πιο συγκεκριμένα, το τριδιάστατο διάνυσμα $\mathbf{v} = (x, y, z)$ θα συμβολίζεται ως

$$\mathbf{v} = (x, y, z) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = [\, x\ y\ z]^T, \tag{2.1}$$

όπου το σύμβολο $[\cdot]^T$ σημαίνει **ανάστροφος** (transpose). Αντίστοιχα θα συμβολίζονται n-διάστατα διανύσματα, για κάθε $n \in \mathbb{N}$.

Κάθε διάνυσμα αντιστοιχεί σε ένα σημείο στον n-διάστατο χώρο. Για παράδειγμα, στο Σχήμα 2.1 σχεδιάζουμε το σημείο $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*)$.

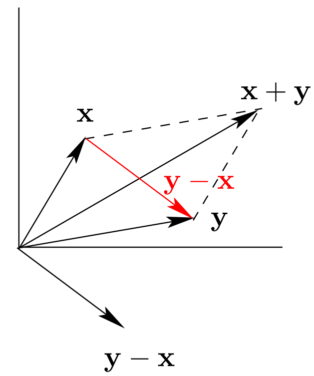
Μία πιο "γεωμετρική" αναπαράσταση ενός διανύσματος, έστω $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*)$, δίδεται από το κατευθυνόμενο ευθύγραμμο τμήμα με αρχή την αρχή των αξόνων, δηλαδή, το σημείο $\mathbf{0} = (0, 0, 0)$, και πέρας το σημείο $\mathbf{x}^*$ (δείτε το Σχήμα 2.2).

2.1.1 Πράξεις διανυσμάτων

Έστω διανύσματα $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ και $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. Το άθροισμά τους $\mathbf{z} = \mathbf{x} + \mathbf{y}$ ορίζεται ως το διάνυσμα $\mathbf{z} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$.

Σχηματικά, το άθροισμα των διανυσμάτων $\mathbf{x}$ και $\mathbf{y}$ αναπαρίσταται από τη διαγώνιο του παραλληλογράμμου με πλευρές τα διανύσματα $\mathbf{x}$ και $\mathbf{y}$ (δείτε το Σχήμα 2.3 για $n = 2$).

Το διάνυσμα $\mathbf{y} - \mathbf{x} = (y_1 - x_1, y_2 - x_2, \ldots, y_n - x_n)$ είναι το διάνυσμα το οποίο θα πρέπει να προσθέσουμε στο $\mathbf{x}$ ώστε να λάβουμε το $\mathbf{y}$ (δείτε το Σχήμα 2.3, για $n = 2$).



Σχήμα 2.3: Πρόσθεση και αφαίρεση διανυσμάτων.

Έστω το διάνυσμα $\mathbf{x} = (x_1, \ldots, x_n)$.

Τότε, το διάνυσμα $a\mathbf{x} = (ax_1, \ldots, ax_n)$, με $a \in \mathbb{R}$, καλείται **κλιμάκωση** του $\mathbf{x}$.

Αν $a > 0$, τότε το $a\mathbf{x}$ έχει την ίδια κατεύθυνση με το $\mathbf{x}$ ενώ, αν $a < 0$, τότε το $a\mathbf{x}$ έχει κατεύθυνση αντίθετη από αυτή του $\mathbf{x}$.

2.1.2 Εσωτερικό γινόμενο και μέτρο

Ορισμός 2.1.2. Έστω τα διανύσματα $\mathbf{x} = (x_1, \ldots, x_n)$ και $\mathbf{y} = (y_1, \ldots, y_n)$. Το εσωτερικό τους γινόμενο ορίζεται ως εξής

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = \sum_{i=1}^{n} x_i y_i. \tag{2.2}$$

Ορισμός 2.1.3. Ως Ευκλείδειο μέτρο του διανύσματος $\mathbf{x} = (x_1, \ldots, x_n)$ ορίζουμε τον μη-αρνητικό πραγματικό αριθμό

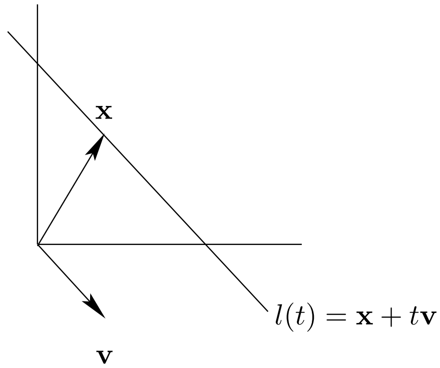
$$\|\mathbf{x}\|_2 = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}} = \left(x_1^2 + \cdots + x_n^2\right)^{\frac{1}{2}}. \tag{2.3}$$

Μπορούν να αποδειχθούν οι παρακάτω σημαντικές ιδιότητες.

1. Αν $\mathbf{x} \in \mathbb{R}^n$, τότε $\|\mathbf{x}\|_2 \geq 0$, με ισότητα αν, και μόνο αν, $\mathbf{x} = \mathbf{0}$.

2. Αν $\mathbf{x} \in \mathbb{R}^n$ και $a \in \mathbb{R}$, τότε $\|a\mathbf{x}\|_2 = |a| \, \|\mathbf{x}\|_2$.

3. Αν $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, τότε $\|\mathbf{x} + \mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2$, με ισότητα αν, και μόνο αν, $\mathbf{x} = c\mathbf{y}$, με $c \in \mathbb{R}$.



Σχήμα 2.4: Ευθεία γραμμή η οποία περνάει από το σημείο $\mathbf{x}$ και είναι παράλληλη με το διάνυσμα $\mathbf{v}$.

Οποιαδήποτε συνάρτηση $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ ικανοποιεί τις παραπάνω συνθήκες καλείται συνάρτηση μέτρου στον $\mathbb{R}^n$.

Θεώρημα 2.1.1. Έστω τα διανύσματα $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Τότε,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \cos \angle(\mathbf{x}, \mathbf{y}), \tag{2.4}$$

όπου $\angle(\mathbf{x}, \mathbf{y})$ είναι η γωνία μεταξύ των διανυσμάτων $\mathbf{x}$ και $\mathbf{y}$.

Απόδειξη. Για $n = 3$, δείτε σελίδα 18 του βιβλίου των Marsden και Tromba. □

Πόρισμα 2.1.1. (*Ανισότητα Cauchy-Schwarz*) Έστω διανύσματα $\mathbf{x}$ και $\mathbf{y}$. Τότε

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \tag{2.5}$$

με ισότητα αν, και μόνο αν, τα $\mathbf{x}$ και $\mathbf{y}$ είναι συγγραμμικά, δηλαδή, $\mathbf{x} = a\mathbf{y}$ για $a \in \mathbb{R}$.

Απόδειξη. Η απόδειξη είναι προφανής από το Θεώρημα 2.1.1 και το ότι $|\cos(\theta)| \leq 1$, για κάθε γωνία θ. □

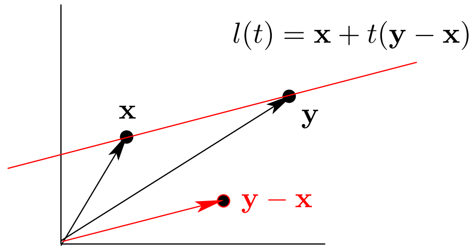
Ορισμός 2.1.4. Τα διανύσματα $\mathbf{x}$ και $\mathbf{y}$ καλούνται ορθογώνια αν $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

Αν τα διανύσματα $\mathbf{x}$ και $\mathbf{y}$ είναι ορθογώνια, τότε γράφουμε $\mathbf{x} \perp \mathbf{y}$.

2.1.3 Ευθείες γραμμές

Έστω τα διανύσματα $\mathbf{x}$ και $\mathbf{v}$. Τα σημεία που περιγράφονται από τη σχέση

$$\mathbf{l}(t) = \mathbf{x} + t\mathbf{v},$$



Σχήμα 2.5: Ευθεία γραμμή η οποία περνάει από τα σημείο $\mathbf{x}$ και $\mathbf{y}$.

για $t \in \mathbb{R}$, ορίζουν την **ευθεία γραμμή** η οποία περνάει από το σημείο $\mathbf{x}$ και είναι παράλληλη με το διάνυσμα $\mathbf{v}$ (δείτε το Σχήμα 2.4).

2.1.4 Ευθεία γραμμή που περνάει από τα σημεία x και y

Έστω τα διανύσματα $\mathbf{x}$ και $\mathbf{y}$. Τα σημεία τα οποία ορίζονται από τη σχέση

$$\mathbf{l}(t) = \mathbf{x} + t(\mathbf{y} - \mathbf{x}), \quad \text{για } t \in \mathbb{R}, \tag{2.6}$$

ορίζουν την ευθεία γραμμή η οποία περνάει από τα σημεία $\mathbf{x}$ και $\mathbf{y}$ (δείτε το Σχήμα 2.5).

Παρατηρούμε ότι, για $t = 0$, $\mathbf{l}(t) = \mathbf{x}$ ενώ, για $t = 1$, $\mathbf{l}(t) = \mathbf{y}$.

Μία εναλλακτική περιγραφή της $\mathbf{l}(t)$ έχει τη μορφή

$$\mathbf{l}(t) = (1 - t)\mathbf{x} + t\mathbf{y}. \tag{2.7}$$

Για $t \in [0, 1]$, η (2.7) ορίζει το ευθύγραμμο τμήμα το οποίο συνδέει τα σημεία $\mathbf{x}$ και $\mathbf{y}$.

2.1.5 Επίπεδο κάθετο σε διάνυσμα

Ως επίπεδο ορίζουμε ένα σύνολο σημείων, έστω $\mathbb{P}$, για το οποίο ισχύει ότι όλα τα σημεία των ευθειών γραμμών οι οποίες συνδέουν δύο οποιαδήποτε σημεία του επιπέδου ανήκουν στο επίπεδο.

Δηλαδή, αν $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{P}$ και $\theta \in \mathbb{R}$, τότε $\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2 \in \mathbb{P}$.

Έστω $\mathbb{P}$ επίπεδο στον $\mathbb{R}^n$, $\mathbf{y}$ σημείο του $\mathbb{P}$, και $\mathbf{a} \in \mathbb{R}^n$ διάνυσμα κάθετο στο $\mathbb{P}$.

Τότε, για κάθε $\mathbf{x} \in \mathbb{P}$, το διάνυσμα $\mathbf{x} - \mathbf{y}$ είναι παράλληλο στο $\mathbb{P}$.

Συνεπώς, το διάνυσμα $\mathbf{a}$ είναι κάθετο στο $\mathbf{x} - \mathbf{y}$.

Η μαθηματική περιγραφή των σημείων του $\mathbb{P}$ έχει ως εξής:

$$\begin{aligned}\mathbb{P} &= \{\mathbf{x} \in \mathbb{R}^n \,|\, \langle \mathbf{x} - \mathbf{y}, \mathbf{a} \rangle = 0\} \\ &= \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x}^T \mathbf{a} = \mathbf{y}^T \mathbf{a}\} \\ &= \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x}^T \mathbf{a} = c\},\end{aligned} \tag{2.8}$$

όπου $c := \mathbf{y}^T \mathbf{a}$.

Γενικά, λέμε ότι η εξίσωση (2.8) ορίζει το **υπερεπίπεδο** (hyperplane) στον $\mathbb{R}^n$ το οποίο περνάει από το σημείο $\mathbf{y}$ και είναι κάθετο στο διάνυσμα $\mathbf{a}$.

Αν $n = 2$, τότε το υπερεπίπεδο είναι απλά μία ευθεία γραμμή, ενώ, αν $n = 3$, τότε το υπερεπίπεδο είναι ένα επίπεδο.

Γενικά, ένα υπερεπίπεδο στον $\mathbb{R}^n$ είναι ένα επίπεδο σύνολο με διάσταση $(n-1)$ (ισοδύναμα, με $(n-1)$ βαθμούς ελευθερίας).

Η εξίσωση (2.8) θα αποδειχθεί εξαιρετικά σημαντική στη συνέχεια.

Άσκηση: Να αποδείξετε ότι το σύνολο $\mathbb{P}$, το οποίο ορίζεται στη σχέση (2.8), είναι επίπεδο.

2.1.6 Ανοιχτά και κλειστά σύνολα

Ορισμός 2.1.5. Έστω $\mathbf{x} \in \mathbb{R}^n$ και $r \in \mathbb{R}_{++}$. Το σύνολο

$$\mathbb{B}(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\|_2 < r\} \tag{2.9}$$

καλείται (Ευκλείδεια) **μπάλα** (ball) με κέντρο το $\mathbf{x}$ και ακτίνα r.

Το σύνολο $\mathbb{B}(\mathbf{x}, r)$ καλείται, επίσης, **γειτονιά** (neighbourhood) του $\mathbf{x}$.

Ορισμός 2.1.6. Το σημείο $\mathbf{x} \in \mathbb{S}$ καλείται **εσωτερικό σημείο** (interior point) του $\mathbb{S}$ αν το $\mathbb{S}$ περιέχει μία γειτονιά του $\mathbf{x}$, δηλαδή, αν υπάρχει $r_{\mathbf{x}} \in \mathbb{R}_{++}$ τέτοιο ώστε $\mathbb{B}(\mathbf{x}, r_{\mathbf{x}}) \subset \mathbb{S}$.

Ορισμός 2.1.7. Το σύνολο των εσωτερικών σημείων του συνόλου $\mathbb{S}$ καλείται **εσωτερικό** (interior) του $\mathbb{S}$.

Ορισμός 2.1.8. Ένα σημείο $\mathbf{x}$ καλείται **συνοριακό σημείο** (boundary point) του συνόλου $\mathbb{S}$ αν *κάθε* γειτονιά του $\mathbf{x}$ περιέχει τουλάχιστον ένα σημείο το οποίο ανήκει στο $\mathbb{S}$ και τουλάχιστον ένα σημείο το οποίο δεν ανήκει στο $\mathbb{S}$. Το σύνολο όλων των συνοριακών σημείων του $\mathbb{S}$ καλείται **σύνορο** (boundary) του $\mathbb{S}$.

Ένα συνοριακό σημείο του συνόλου $\mathbb{S}$ μπορεί να μην ανήκει στο $\mathbb{S}$.

Ορισμός 2.1.9. Ένα σύνολο καλείται **ανοιχτό** (open) αν περιέχει μόνο τα εσωτερικά του σημεία ή, ισοδύναμα, αν δεν περιέχει κανένα από τα συνοριακά του σημεία.

Ορισμός 2.1.10. Ένα σύνολο καλείται **κλειστό** (closed) αν περιέχει το σύνορό του.

Μπορεί να αποδειχθεί ότι ένα σύνολο είναι ανοιχτό αν, και μόνο αν, το συμπλήρωμά του είναι κλειστό.

Ένα σύνολο μπορεί να μην είναι ούτε ανοιχτό ούτε κλειστό (να δώσετε παραδείγματα).

Ορισμός 2.1.11. Ένα σύνολο καλείται **φραγμένο** (bounded) αν περιέχεται σε μία μπάλα με πεπερασμένη ακτίνα.

Ορισμός 2.1.12. Ένα σύνολο καλείται **συμπαγές** (compact) αν είναι κλειστό και φραγμένο.

Θεώρημα 2.1.2. (*Θεώρημα Weierstrass*) Έστω $\mathbb{U} \subseteq \mathbb{R}^n$ συμπαγές σύνολο και $f : \mathbb{U} \to \mathbb{R}$ συνεχής συνάρτηση. Τότε, υπάρχει σημείο $\mathbf{x}_0 \in \mathbb{U}$ τέτοιο ώστε $f(\mathbf{x}_0) \leq f(\mathbf{x})$, για κάθε $\mathbf{x} \in \mathbb{U}$.

Απόδειξη. Η απόδειξη απαιτεί τη χρήση κάποιων προχωρημένων εννοιών Πραγματικής Ανάλυσης και είναι εκτός των πλαισίων του μαθήματος. □

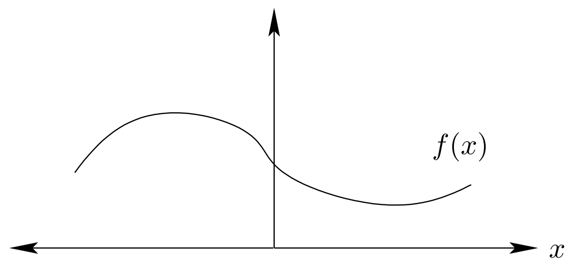
Με λόγια, το σημαντικό αυτό θεώρημα λέει ότι, αν το $\mathbb{U}$ είναι συμπαγές και η f συνεχής, τότε υπάρχει σημείο του $\mathbb{U}$ το οποίο επιτυγχάνει την ελάχιστη τιμή της f στο $\mathbb{U}$.

Αν το $\mathbb{U}$ δεν είναι κλειστό ή φραγμένο, τότε μπορούμε εύκολα να κατασκευάσουμε παραδείγματα στα οποία η ελάχιστη τιμή της f στο $\mathbb{U}$ δεν ορίζεται.

2.2 Στοιχεία θεωρίας συναρτήσεων

2.2.1 Συναρτήσεις

Ορισμός 2.2.1. Έστω $\mathbb{U} \subseteq \mathbb{R}^n$. Η απεικόνιση $f : \mathbb{U} \to \mathbb{R}^m$, η οποία αντιστοιχεί κάθε στοιχείο $\mathbf{x} \in \mathbb{U}$ σε ένα και μόνο ένα στοιχείο του $\mathbb{R}^m$, καλείται **συνάρτηση** (function) από



Σχήμα 2.6: Γράφημα πραγματικής συνάρτησης μίας μεταβλητής.

το $\mathbb{U}$ στο $\mathbb{R}^m$.

Αν στον ορισμό 2.2.1 έχουμε ότι $m = 1$, τότε η $\mathbf{f}$ καλείται πραγματική συνάρτηση n μεταβλητών και συμβολίζεται ως $f(\mathbf{x})$ ή ως $f(x_1, \dots, x_n)$.

Για παράδειγμα, η

$$f(x_1, x_2) = \sqrt{x_1^2 + x_2^2}, \text{ με } x_1, x_2 \in \mathbb{R},$$

είναι πραγματική συνάρτηση δύο μεταβλητών, ενώ η

$$f(x_1, x_2, x_3) = x_1(x_2^2 + \log_2 x_3), \text{ με } x_1, x_2 \in \mathbb{R}, x_3 \in \mathbb{R}_{++},$$

είναι πραγματική συνάρτηση τριών μεταβλητών.

Ορισμός 2.2.2. Έστω $f : \mathbb{U} \subseteq \mathbb{R}^n \to \mathbb{R}$. Τότε, το σύνολο

$$\mathbb{G}^f = \{(\mathbf{x}, f(\mathbf{x})) \in \mathbb{R}^{n+1} \,|\, \mathbf{x} \in \mathbb{U}\} \tag{2.10}$$

καλείται **γράφημα** (graph) της f.

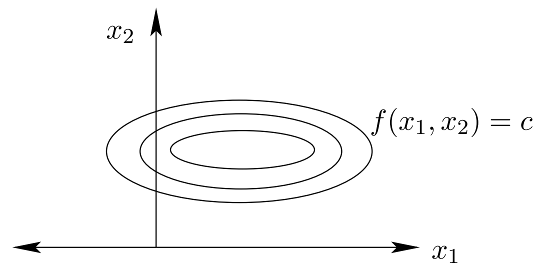
Αν $n = 1$, τότε το γράφημα είναι μία καμπύλη στον $\mathbb{R}^2$ (δείτε το Σχήμα 2.6), ενώ, αν $n = 2$, τότε το γράφημα είναι μία επιφάνεια στον $\mathbb{R}^3$.

Αν $n > 2$, τότε το γράφημα είναι μία υπερεπιφάνεια στον $\mathbb{R}^{n+1}$, η οποία δεν μπορεί να οπτικοποιηθεί.

2.2.2 Καμπύλες

Ορισμός 2.2.3. Κάθε συνάρτηση $\mathbf{f} : \mathbb{A} \subseteq \mathbb{R} \to \mathbb{R}^n$ καλείται **καμπύλη** (curve) στον $\mathbb{R}^n$.

Παράδειγμα 2.2.1. Η συνάρτηση $\mathbf{f} : [0, 2\pi] \to \mathbb{R}^2$, με $\mathbf{f}(t) = (\cos(t), \sin(t))$ ορίζει τον κύκλο με κέντρο το σημείο $(0, 0)$ και ακτίνα 1.



Σχήμα 2.7: Σύνολα στάθμης πραγματικής συνάρτησης δύο μεταβλητών.

2.2.3 Σύνολα στάθμης

Ορισμός 2.2.4. Έστω $f : \mathbb{U} \subseteq \mathbb{R}^n \to \mathbb{R}$ και $c \in \mathbb{R}$. Το σύνολο

$$\mathbb{S}_c := \{\mathbf{x} \in \mathbb{U} \,|\, f(\mathbf{x}) = c\} \tag{2.11}$$

καλείται **σύνολο στάθμης** (level set) της f, για την τιμή c.

Αν $n = 2$, τότε μιλάμε για καμπύλη στάθμης (δείτε το Σχήμα 2.7), ενώ αν $n = 3$ τότε μιλάμε για επιφάνεια στάθμης.

2.2.4 Όρια ακολουθιών - Συνέχεια συναρτήσεων

Υποθέτουμε ότι ο αναγνώστης είναι εξοικειωμένος με τις έννοιες των ακολουθιών διανυσμάτων και σύγκλισης ακολουθιών διανυσμάτων καθώς και με την έννοια της συνέχειας συναρτήσεων. Το βιβλίο των Marsden και Tromba περιέχει εκτενή κάλυψη αυτών των βασικών εννοιών.

2.2.5 Παράγωγος

Η έννοια της παραγώγισης έχει να κάνει με την προσέγγιση συναρτήσεων από γραμμικές (για την ακρίβεια, αφφινικές) συναρτήσεις. Στη συνέχεια, παραθέτουμε βασικούς ορισμούς και ιδιότητες των παραγώγων συναρτήσεων.

Ορισμός 2.2.5. Έστω $\mathbb{U}$ ανοιχτό υποσύνολο του $\mathbb{R}^n$ και $f : \mathbb{U} \to \mathbb{R}^m$. Η $\mathbf{f}$ καλείται **διαφορίσιμη** (differentiable) στο σημείο $\mathbf{x} \in \mathbb{U}$ αν υπάρχει $m \times n$ πίνακας $\mathbf{A}^f_{\mathbf{x}}$ τέτοιος ώστε

$$\lim_{\mathbf{y} \to \mathbf{x}} \left(\frac{\|f(\mathbf{y}) - f(\mathbf{x}) - \mathbf{A}^f_{\mathbf{x}}(\mathbf{y} - \mathbf{x})\|}{\|\mathbf{y} - \mathbf{x}\|} \right) = 0. \tag{2.12}$$

Ο πίνακας $\mathbf{A}^f_{\mathbf{x}}$ καλείται **παράγωγος** (derivative) της f στο σημείο $\mathbf{x}$ και συμβολίζεται ως $Df(\mathbf{x})$.

Στη συνέχεια, αναφέρουμε ως θεωρήματα τρεις σημαντικές ιδιότητες της παραγώγου.

Θεώρημα 2.2.1. Αν $f : \mathbb{R}^n \to \mathbb{R}^m$ και $g : \mathbb{R}^n \to \mathbb{R}^m$ είναι διαφορίσιμες στο σημείο $\mathbf{x} \in \mathbb{R}^n$, τότε και η $(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ είναι διαφορίσιμη στο $\mathbf{x}$ και

$$D(f+g)(\mathbf{x}) = Df(\mathbf{x}) + Dg(\mathbf{x}). \tag{2.13}$$

Απόδειξη. Βλέπε σελίδα 101 του βιβλίου των Marsden και Tromba. □

Θεώρημα 2.2.2. Αν $f : \mathbb{R}^n \to \mathbb{R}$ και $g : \mathbb{R}^n \to \mathbb{R}$ είναι διαφορίσιμες στο σημείο $\mathbf{x} \in \mathbb{R}^n$, τότε και η $(f \cdot g)(\mathbf{x}) = f(\mathbf{x}) \cdot g(\mathbf{x})$ είναι διαφορίσιμη στο $\mathbf{x}$ και

$$D(f \cdot g)(\mathbf{x}) = g(\mathbf{x})Df(\mathbf{x}) + f(\mathbf{x})Dg(\mathbf{x}). \tag{2.14}$$

Απόδειξη. Βλέπε σελίδα 101 του βιβλίου των Marsden και Tromba. □

Θεώρημα 2.2.3. Έστω $f : \mathbb{R}^n \to \mathbb{R}^m$ και $h : \mathbb{R}^k \to \mathbb{R}^n$. Αν η h είναι διαφορίσιμη στο $\mathbf{x} \in \mathbb{R}^k$ και η f είναι διαφορίσιμη στο $h(\mathbf{x}) \in \mathbb{R}^n$, τότε η $f \circ h$ είναι διαφορίσιμη στο $\mathbf{x} \in \mathbb{R}^k$ και

$$D(f \circ h)(\mathbf{x}) = Df(h(\mathbf{x}))Dh(\mathbf{x}), \tag{2.15}$$

όπου στο δεξιό μέλος της (2.15) έχουμε πολλαπλασιασμό πινάκων.

Απόδειξη. Βλέπε σελίδα 102 του βιβλίου των Marsden και Tromba. □

Στο Παράρτημα στο τέλος του Κεφαλαίου, αποδεικνύουμε αναλυτικά μία ειδική περίπτωση του Θεωρήματος 2.2.3.

Αν μία συνάρτηση f είναι διαφορίσιμη και η παράγωγός της, Df, είναι συνεχής συνάρτηση, τότε λέμε ότι η f ανήκει στην οικογένεια συναρτήσεων C^1.

Ορισμός 2.2.6. Έστω $\mathbb{U}$ ανοιχτό υποσύνολο του $\mathbb{R}^n$, $f : \mathbb{U} \to \mathbb{R}$, $\mathbf{x} \in \mathbb{U}$, και $\mathbf{e}_j$, για $j = 1, \ldots, n$, το n-διάστατο διάνυσμα με στοιχεία παντού 0 εκτός από την j-οστή θέση στην οποία έχει στοιχείο 1. Αν υπάρχει το όριο

$$\lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h}, \tag{2.16}$$

τότε καλείται **μερική παράγωγος** (partial derivative) της f ως προς την j-οστή μεταβλητή, x_j, στη θέση $\mathbf{x}$, και συμβολίζεται ως $\frac{\partial f}{\partial x_j}(\mathbf{x})$ ή ως $\frac{\partial f}{\partial x_j}(x_1, \ldots, x_n)$.

Στη συνέχεια, παραθέτουμε ένα σημαντικό θεώρημα το οποίο συνδέει την παράγωγο μίας συνάρτησης f με τις μερικές παραγώγους της.

Θεώρημα 2.2.4. Έστω $\mathbb{U} \subseteq \mathbb{R}^n$ ανοιχτό σύνολο και $f : \mathbb{U} \to \mathbb{R}$.

1. Αν η f είναι διαφορίσιμη στο $\mathbf{x} \in \mathbb{U}$, τότε οι μερικές παράγωγοι $\frac{\partial f}{\partial x_i}(\mathbf{x})$ υπάρχουν και $Df(\mathbf{x}) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1} \ \cdots \ \frac{\partial f(\mathbf{x})}{\partial x_n}\right]$.

2. Αν οι μερικές παράγωγοι της f στο σημείο $\mathbf{x}$ υπάρχουν και είναι συνεχείς, τότε η f είναι διαφορίσιμη στο $\mathbf{x}$ και $Df(\mathbf{x}) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1} \ \cdots \ \frac{\partial f(\mathbf{x})}{\partial x_n}\right]$.

3. Η f είναι C^1 στο $\mathbb{U}$ αν, και μόνο αν, οι μερικές παράγωγοι της f υπάρχουν και είναι συνεχείς στο $\mathbb{U}$.

Απόδειξη. Δείτε ένα καλό βιβλίο Λογισμού. □

Συνεπώς, η ύπαρξη των μερικών παραγώγων της f δεν συνεπάγεται αυτόματα τη διαφορισιμότητα της f. Η συνέχεια των μερικών παραγώγων είναι εξαιρετικά σημαντική.

Ορισμός 2.2.7. Αν η $f : \mathbb{R}^n \to \mathbb{R}$ είναι διαφορίσιμη, τότε η συνάρτηση $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$, η οποία ορίζεται ως

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix} = Df(\mathbf{x})^T, \tag{2.17}$$

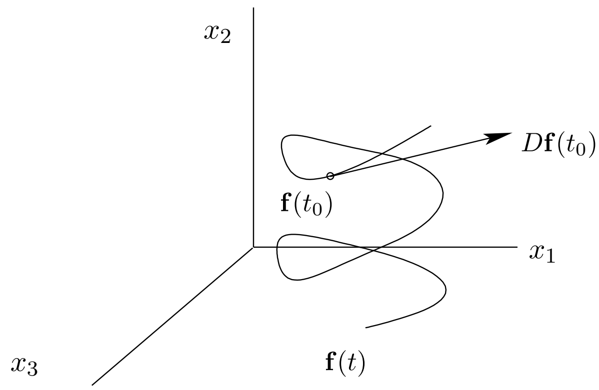
καλείται **βαθμίδα** (gradient) της f στο σημείο $\mathbf{x}$.

Αργότερα, θα αποδείξουμε κάποιες εξαιρετικά σημαντικές ιδιότητες της $\nabla f(\mathbf{x})$.

Ορισμός 2.2.8. Έστω $f : \mathbb{R}^n \to \mathbb{R}$. Αν η ∇f είναι διαφορίσιμη, τότε λέμε ότι η f είναι διπλά διαφορίσιμη. Η δεύτερη παράγωγος της f είναι η παράγωγος της ∇f και συμβολίζεται ως

$$D^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}. \tag{2.18}$$

Ο $(n \times n)$ πίνακας $D^2 f$ καλείται και **Εσσιανή** (Hessian) της f.



Σχήμα 2.8: Παράγωγος καμπύλης.

2.2.6 Παράγωγος καμπύλης

Ορισμός 2.2.9. Αν $\mathbf{f} : \mathbb{A} \subseteq \mathbb{R} \to \mathbb{R}^n$, με $\mathbf{f}(t) = (f_1(t), \ldots, f_n(t))$, τότε το $(n \times 1)$ διάνυσμα $D\mathbf{f}(t_0) = (Df_1(t_0), \ldots, Df_n(t_0))$ καλείται παράγωγος της καμπύλης $\mathbf{f}$ στο σημείο $(f_1(t_0), \ldots, f_n(t_0))$.

Η παράγωγος $D\mathbf{f}$ μπορεί να εκφραστεί ως

$$D\mathbf{f}(t_0) = \lim_{h \to 0} \frac{\mathbf{f}(t_0 + h) - \mathbf{f}(t_0)}{h}. \tag{2.19}$$

Το διάνυσμα $D\mathbf{f}(t_0)$ είναι παράλληλο στην ευθεία η οποία εφάπτεται στην καμπύλη $\mathbf{f}$ στο σημείο $\mathbf{f}(t_0)$ (δείτε το Σχήμα 2.8) και εκφράζεται ως εξής:

$$\mathbf{l}(t) = \mathbf{f}(t_0) + tD\mathbf{f}(t_0). \tag{2.20}$$

2.2.7 Παράγωγος κατά κατεύθυνση

Ορισμός 2.2.10. Έστω $\mathbb{U} \subseteq \mathbb{R}^n$ ανοιχτό σύνολο και $f : \mathbb{U} \to \mathbb{R}$. Έστω $\mathbf{x} \in \mathbb{U}$ και $\mathbf{0} \neq \mathbf{h} \in \mathbb{R}^n$. Αν το όριο

$$\lim_{t \to 0^+} \left(\frac{f(\mathbf{x} + t\mathbf{h}) - f(\mathbf{x})}{t} \right). \tag{2.21}$$

υπάρχει, καλείται **παράγωγος κατά κατεύθυνση** (directional derivative) της f, στο σημείο $\mathbf{x}$ και στην κατεύθυνση $\mathbf{h}$, και συμβολίζεται ως $Df(\mathbf{x}; \mathbf{h})$.

Αν η συνάρτηση f είναι διαφορίσιμη, τότε μία ισοδύναμη έκφραση για την παράγωγο κατά κατεύθυνση έχει ως εξής

$$Df(\mathbf{x}; \mathbf{h}) = \frac{d}{dt} f(\mathbf{x} + t\mathbf{h}) \bigg|_{t=0}. \tag{2.22}$$

Αυτό συμβαίνει διότι

$$\begin{aligned} \frac{d}{dt} f(\mathbf{x} + t\mathbf{h}) \bigg|_{t=0} &= \lim_{\Delta t \to 0} \frac{f(\mathbf{x} + (t + \Delta t)\mathbf{h}) - f(\mathbf{x} + t\mathbf{h})}{\Delta t} \bigg|_{t=0} \\ &= \lim_{\Delta t \to 0} \frac{f(\mathbf{x} + \Delta t\, \mathbf{h}) - f(\mathbf{x})}{\Delta t}. \end{aligned} \tag{2.23}$$

Συνεπώς, στην περίπτωση αυτή, υπάρχει το αμφίπλευρο όριο, $\lim_{t \to 0}$, και ισούται με το όριο από τα δεξιά, $\lim_{t \to 0^+}$.

Θεώρημα 2.2.5. Έστω $f : \mathbb{R}^n \to \mathbb{R}$ διαφορίσιμη συνάρτηση και $\mathbf{x} \in \mathbb{R}^n$. Τότε, για κάθε $\mathbf{h} \in \mathbb{R}^n$, η παράγωγος κατά κατεύθυνση $Df(\mathbf{x}; \mathbf{h})$ υπάρχει και δίνεται από τη σχέση

$$Df(\mathbf{x}; \mathbf{h}) = Df(\mathbf{x})\, \mathbf{h} = \nabla f(\mathbf{x})^T \mathbf{h}. \tag{2.24}$$

Απόδειξη. Έστω η $\mathbf{c}(t) = \mathbf{x} + t\mathbf{h}$. Τότε $f(\mathbf{x} + t\mathbf{h}) = f(\mathbf{c}(t))$. Από τον κανόνα της αλυσίδας, έχουμε ότι

$$\frac{d}{dt} f(\mathbf{c}(t)) = Df(\mathbf{c}(t)) D\mathbf{c}(t). \tag{2.25}$$

Επιπλέον, $\mathbf{c}(0) = \mathbf{x}$ και $D\mathbf{c}(t) = \mathbf{h}$. Συνεπώς,

$$\begin{aligned} Df(\mathbf{x}; \mathbf{h}) = \frac{d}{dt} f(\mathbf{x} + t\mathbf{h}) \bigg|_{t=0} &= \frac{d}{dt} f(\mathbf{c}(t)) \bigg|_{t=0} = Df(\mathbf{c}(t)) D\mathbf{c}(t)\big|_{t=0} \\ &= Df(\mathbf{c}(0)) D\mathbf{c}(0) = Df(\mathbf{x})\mathbf{h} = \nabla f(\mathbf{x})^T \mathbf{h}, \end{aligned} \tag{2.26}$$

ολοκληρώνοντας την απόδειξη. □

Επιλέγοντας $\mathbf{h} \in \mathbb{R}^n$ με $\|\mathbf{h}\|_2 = 1$, μπορούμε να ερμηνεύουμε την $Df(\mathbf{x}; \mathbf{h})$ σαν τον ρυθμό μεταβολής της f στο σημείο $\mathbf{x}$ και στην κατεύθυνση $\mathbf{h}$.

Μία άμεση συνέπεια του Θεωρήματος 2.2.5 έχει ως εξής.

Πόρισμα 2.2.1. Αν η $Df(\mathbf{x}; \mathbf{h})$ υπάρχει, τότε υπάρχει και η $Df(\mathbf{x}; -\mathbf{h})$ και

$$Df(\mathbf{x}; -\mathbf{h}) = -Df(\mathbf{x}; \mathbf{h}). \tag{2.27}$$

Ένα εξαιρετικά σημαντικό θεώρημα σχετικό με το ρυθμό μεταβολής μίας συνάρτησης f έχει ως εξής.

Θεώρημα 2.2.6. Έστω $f : \mathbb{R}^n \to \mathbb{R}$ λεία πραγματική συνάρτηση και $\mathbf{x}_0 \in \mathbb{R}^n$. Έστω $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$. Τότε, το διάνυσμα $\nabla f(\mathbf{x}_0)$ "δείχνει" προς εκείνη την κατεύθυνση κατά μήκος της οποίας ο ρυθμός αύξησης της f στο $\mathbf{x}_0$ είναι μέγιστος, ενώ το $-\nabla f(\mathbf{x}_0)$ "δείχνει" προς εκείνη την κατεύθυνση κατά μήκος της οποίας ο ρυθμός μείωσης της f στο $\mathbf{x}_0$ είναι μέγιστος.

Απόδειξη. Έστω $\mathbf{h}$ διάνυσμα με μοναδιαίο Ευκλείδειο μέτρο στον $\mathbb{R}^n$. Ο ρυθμός μεταβολής της f στο σημείο $\mathbf{x}_0$ και στην κατεύθυνση $\mathbf{h}$ ισούται με $\nabla f(\mathbf{x}_0)^T\mathbf{h}$. Όμως

$$\nabla f(\mathbf{x}_0)^T\mathbf{h} = \|\nabla f(\mathbf{x}_0)\|_2 \cos\angle\left(\nabla f(\mathbf{x}_0), \mathbf{h}\right) \tag{2.28}$$

και

$$\left|\nabla f(\mathbf{x}_0)^T\mathbf{h}\right| = \|\nabla f(\mathbf{x}_0)\|_2 \left|\cos\angle\left(\nabla f(\mathbf{x}_0), \mathbf{h}\right)\right| \leq \|\nabla f(\mathbf{x}_0)\|_2 \tag{2.29}$$

από το οποίο λαμβάνουμε ότι

$$-\|\nabla f(\mathbf{x}_0)\|_2 \leq \nabla f(\mathbf{x}_0)^T\mathbf{h} \leq \|\nabla f(\mathbf{x}_0)\|_2 \tag{2.30}$$

Η αριστερή ανισότητα επιτυγχάνεται για $\mathbf{h} = -\frac{1}{\|\nabla f(\mathbf{x}_0)\|_2}\nabla f(\mathbf{x}_0)$ ενώ η δεξιά ανισότητα επιτυγχάνεται για $\mathbf{h} = \frac{1}{\|\nabla f(\mathbf{x}_0)\|_2}\nabla f(\mathbf{x}_0)$. □

2.2.8 Βαθμίδα και σύνολα στάθμης συνάρτησης

Σε αυτό το εδάφιο, αποδεικνύουμε ένα ακόμα σημαντικό αποτέλεσμα για τη βαθμίδα συνάρτησης.

Επαναλαμβάνουμε ότι μία συνάρτηση $\mathbf{c} : [a, b] \subseteq \mathbb{R} \to \mathbb{R}^m$ καλείται καμπύλη. Η καμπύλη $\mathbf{c}(t)$ μπορεί να εκφραστεί ως

$$\mathbf{c}(t) = \begin{bmatrix} c_1(t) \\ \vdots \\ c_m(t) \end{bmatrix} \tag{2.31}$$

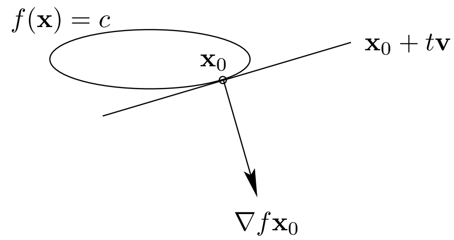
με $c_i : [a, b] \to \mathbb{R}$, για $i = 1, \ldots, m$.

Αν η $\mathbf{c}$ είναι παραγωγίσιμη στο $t \in [a, b]$, τότε η παράγωγός της είναι το m-διάστατο διάνυσμα το οποίο ορίζεται ως

$$D\mathbf{c}(t) = \lim_{h \to 0} \frac{\mathbf{c}(t + h) - \mathbf{c}(t)}{h}. \tag{2.32}$$

Μπορεί να αποδειχεί ότι το i-οστό στοιχείο του $D\mathbf{c}(t)$ ισούται με την παράγωγο της c_i, για $i = 1, \ldots, m$.

Επιπλέον, το $D\mathbf{c}(t)$ ορίζει το διάνυσμα το οποίο εφάπτεται στην καμπύλη $\mathbf{c}$ στο σημείο $\mathbf{c}(t)$.



Σχήμα 2.9: Σχέση διανύσματος βαθμίδας ∇f και ισοϋψούς επιφάνειας της f στο σημείο $\mathbf{x}_0$.

Θεώρημα 2.2.7. Έστω $f : \mathbb{R}^n \to \mathbb{R}$ συνεχώς διαφορίσιμη πραγματική συνάρτηση και $\mathbf{x}_0$ σημείο του συνόλου στάθμης

$$\mathbb{S}_c = \{\mathbf{x} \in \mathbb{R}^n \,|\, f(\mathbf{x}) = c\}\,.$$

Τότε, το διάνυσμα βαθμίδας $\nabla f(\mathbf{x}_0)$ είναι ορθογώνιο στην επιφάνεια στάθμης με την εξής έννοια (δείτε το Σχήμα 2.9). Αν $\mathbf{c}(t)$ είναι λεία καμπύλη η οποία περιέχεται στο σύνολο $\mathbb{S}_c$, με $\mathbf{c}(0) = \mathbf{x}_0$, και $\mathbf{v}$ το εφαπτόμενο διάνυσμα στην $\mathbf{c}(t)$, για $t = 0$, τότε

$$\nabla f(\mathbf{x}_0)^T \mathbf{v} = 0. \tag{2.33}$$

Απόδειξη. Αφού η καμπύλη $\mathbf{c}(t)$ περιέχεται στο σύνολο $\mathbb{S}_c$, έχουμε ότι $f(\mathbf{c}(t)) = c$.

Επιπλέον, $\mathbf{c}(0) = \mathbf{x}_0$ και $D\mathbf{c}(0) = \mathbf{v}$.

Από τον κανόνα αλυσίδας, έχουμε ότι

$$\begin{aligned} 0 = \left.\frac{d}{dt} f(\mathbf{c}(t))\right|_{t=0} &= Df(\mathbf{c}(t)) D\mathbf{c}(t)|_{t=0} \\ &= Df(\mathbf{c}(0)) D\mathbf{c}(0) \\ &= Df(\mathbf{x}_0)\mathbf{v} \\ &= \nabla f(\mathbf{x}_0)^T \mathbf{v}, \end{aligned} \tag{2.34}$$

ολοκληρώνοντας την απόδειξη. □

Ως απόρροια του Θεωρήματος 2.2.7, είναι λογικό να ορίσουμε το επίπεδο το οποίο εφάπτεται στην ισοϋψή επιφάνεια της f, $\mathbb{S}_c$, στο σημείο $\mathbf{x}_0 \in \mathbb{S}_c$, ως εξής.

Ορισμός 2.2.11. Έστω $f : \mathbb{R}^n \to \mathbb{R}$ διαφορίσιμη και $\mathbb{S}_c = \{\mathbf{x} \in \mathbb{R}^n \,|\, f(\mathbf{x}) = c\}$. Τότε, το επίπεδο το οποίο εφάπτεται στην $\mathbb{S}_c$ στο σημείο $\mathbf{x}_0 \in \mathbb{S}_c$ ορίζεται από την εξίσωση

$$\mathbb{P}^f_{\mathbf{x}_0} = \{\mathbf{x} \in \mathbb{R}^n \,|\, \nabla f(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0) = 0\}. \tag{2.35}$$

Προφανώς, το $\mathbb{P}^f_{\mathbf{x}_0}$ περιέχει το σημείο $\mathbf{x}_0$.

Αν το $\mathbf{x} \neq \mathbf{x}_0$ ανήκει στο εφαπτόμενο επίπεδο, τότε το διάνυσμα $\mathbf{x} - \mathbf{x}_0$ είναι παράλληλο στο εφαπτόμενο επίπεδο.

Όμως, έχουμε αποδείξει ότι το $\nabla f(\mathbf{x}_0)$ είναι κάθετο σε κάθε εφαπτόμενο διάνυσμα καμπύλης η οποία ανήκει στην επιφάνεια $\mathbb{S}_c$ και περνάει από το σημείο $\mathbf{x}_0$.

Συνεπώς, ο ορισμός (2.35) είναι ικανοποιητικός.

2.2.9 Θεώρημα Taylor

Έστω $f : \mathbb{R}^n \to \mathbb{R}$ λεία πραγματική συνάρτηση. Τότε ισχύουν οι παρακάτω εξαιρετικά σημαντικές σχέσεις

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + O(\|\mathbf{x} - \mathbf{x}_0\|^2) \\ &= f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) \\ &\qquad + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T D^2 f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + O(\|\mathbf{x} - \mathbf{x}_0\|^3), \end{aligned}$$

τις οποίες καλούμε αναπτύγματα Taylor πρώτης και δεύτερης τάξης, αντίστοιχα.

Αν στις παραπάνω εκφράσεις αγνοήσουμε τους όρους οι οποίοι συμβολίζονται με $O(\cdot)$, τότε έχουμε τις προσεγγίσεις Taylor πρώτης και δεύτερης τάξης, αντίστοιχα.

Επιπλέον, μπορεί να αποδειχθεί ότι

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{z})^T(\mathbf{x} - \mathbf{x}_0) \tag{2.36}$$

για κάποιο $\mathbf{z} = \theta\mathbf{x} + (1-\theta)\mathbf{x}_0$, με $0 \leq \theta \leq 1$, και

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T D^2 f(\mathbf{w})(\mathbf{x} - \mathbf{x}_0), \tag{2.37}$$

για κάποιο $\mathbf{w} = \theta\mathbf{x} + (1-\theta)\mathbf{x}_0$, με $0 \leq \theta \leq 1$.

2.3 Παράρτημα

Σε αυτό το εδάφιο, χρησιμοποιούμε αναπτύγματα Taylor πρώτης τάξης και δίνουμε μία απόδειξη μίας ειδικής μορφής του κανόνα αλυσίδας.

Έστω $\mathbf{c} : \mathbb{R} \to \mathbb{R}^n$ λεία καμπύλη και $f : \mathbb{R}^n \to \mathbb{R}$ λεία πραγματική συνάρτηση. Ορίζουμε $g : \mathbb{R} \to \mathbb{R}$, με $g(t) = (f \circ \mathbf{c})(t) = f(\mathbf{c}(t))$. Έστω ότι θέλουμε να υπολογίσουμε την παράγωγο $Dg(t) = D(f \circ \mathbf{c})(t)$. Υπενθυμίζουμε ότι

$$\mathbf{c}(t + \Delta t) = \mathbf{c}(t) + D\mathbf{c}(t)\Delta t + O(\Delta t^2), \tag{2.38}$$

$$f(\mathbf{x} + \Delta \mathbf{x}) = f(\mathbf{x}) + Df(\mathbf{x})\Delta \mathbf{x} + O(\|\Delta \mathbf{x}\|_2^2) \tag{2.39}$$

Τότε

$$\begin{aligned} Dg(t) &= \lim_{\Delta t \to 0} \frac{g(t + \Delta t) - g(t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{f(\mathbf{c}(t + \Delta t)) - f(\mathbf{c}(t))}{\Delta t} \\ &\overset{(2.38)}{=} \lim_{\Delta t \to 0} \frac{f\left(\mathbf{c}(t) + D\mathbf{c}(t)\Delta t + O(\Delta t^2)\right) - f(\mathbf{c}(t))}{\Delta t} \\ &\overset{(2.39)}{=} \lim_{\Delta t \to 0} \frac{f\left(\mathbf{c}(t)\right) + Df(\mathbf{c}(t))D\mathbf{c}(t)\Delta t + O(\Delta t^2) - f(\mathbf{c}(t))}{\Delta t} \\ &= Df(\mathbf{c}(t))D\mathbf{c}(t). \end{aligned} \tag{2.40}$$

2.4 Σύντομη επισκόπηση τεχνικών αποδείξεων

Προτάσεις (statements) συμβολίζονται με κεφαλαία γράμματα, π.χ., P, Q, και θα είναι αληθείς ή ψευδείς. Για παράδειγμα, η πρόταση

$$P = \text{“Το 2 είναι άρτιος αριθμός.”}$$

είναι αληθής.

Αν P είναι πρόταση, τότε η **άρνηση** (negation) της P είναι πρόταση, συμβολίζεται με not P, και είναι αληθής αν η P είναι ψευδής και ψευδής εάν η P είναι αληθής (δείτε τον παρακάτω πίνακα αληθείας).

P	not P
T	F
F	T

Ένας σημαντικός τρόπος κατασκευής προτάσεων είναι μέσω συνεπαγωγής. Συμβολικά, έχουμε

$$\begin{gathered}(P \Rightarrow Q) \qquad (\text{Αν } P, \text{ τότε } Q) \\ (P \text{ συνεπάγεται } Q), \qquad (Q, \text{ μόνο αν } P).\end{gathered} \tag{2.41}$$

Η πρόταση P καλείται **υπόθεση** (hypothesis) και η πρόταση Q καλείται **συμπέρασμα** (conclusion).

Αν ισχύει η συνεπαγωγή, θα λέμε ότι η P είναι ικανή για την Q και η Q είναι αναγκαία για την P.

Ο πίνακας αληθείας της πρότασης $P \Rightarrow Q$ έχει ως εξής

P	Q	$P \Rightarrow Q$
F	F	T
F	T	T
T	F	F
T	T	T

Δηλαδή, αν η υπόθεση P είναι ψευδής, τότε η συνεπαγωγή είναι αληθής ανεξάρτητα από το αν το συμπέρασμα Q είναι αληθές ή ψευδές (αυτό στην αρχή ίσως φαίνεται λίγο περίεργο). Αν η υπόθεση P είναι αληθής και το συμπέρασμα Q ψευδές, τότε η συνεπαγωγή είναι ψευδής, ενώ αν και η υπόθεση και το συμπέρασμα είναι αληθή, τότε η συνεπαγωγή είναι αληθής.

Η πρόταση $Q \Rightarrow P$ καλείται **αντίστροφη** (inverse) της $P \Rightarrow Q$ ενώ η πρόταση $(\text{not}\, Q) \Rightarrow (\text{not}\, P)$ καλείται **αντιθετοαντίστροφη** (contrapositive) της $P \Rightarrow Q$. Οι αντίστοιχοι πίνακες αληθείας έχουν ως εξής:

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$(\text{not}\, Q) \Rightarrow (\text{not}\, P)$	$\text{not}(P \wedge \text{not}(Q))$
F	F	T	T	T	T
F	T	T	F	T	T
T	F	F	T	F	F
T	T	T	T	T	T

Παρατηρούμε ότι η πρόταση $P \Rightarrow Q$ είναι λογικά ισοδύναμη (έχει τον ίδιο πίνακα αληθείας) με την πρόταση $\text{not}\, Q \Rightarrow \text{not}\, P$. Μία άλλη πρόταση ισοδύναμη με την $P \Rightarrow Q$ είναι η $\text{not}\,(P\, \text{and}\, (\text{not}\, Q))$.

Για να αποδείξουμε μία πρόταση της μορφής $P \Rightarrow Q$ έχουμε τις εξής βασικές επιλογές:

1. Ευθεία απόδειξη: Εκκινούμε από την υπόθεση P και μετά από "σωστούς συλλογισμούς" καταλήγουμε στο συμπέρασμα Q.

2. Απόδειξη μέσω αντιθετοαντιστροφής: Εκκινούμε από υπόθεση $(\text{not}\, Q)$ και μετά από "σωστούς συλλογισμούς" καταλήγουμε στο συμπέρασμα $(\text{not}\, P)$.

3. Απόδειξη μέσω εις άτοπον απαγωγής: Εκκινούμε από υπόθεση $P\, \text{and}\, (\text{not}\, Q)$ και μετά από "σωστούς συλλογισμούς" καταλήγουμε σε κάτι ψευδές.

2.4.1 Quantifiers

Σημαντικά σύμβολα τα οποία χρησιμοποιούμε για την κατασκευή προτάσεων είναι τα εξής: $\forall$ και $\exists$. Το πρώτο σύμβολο δηλώνει τον universal quantifier ενώ το δεύτερο τον existential quantifier.

Για παράδειγμα, θεωρήστε τις προτάσεις

$$(\forall x)(\exists y)(x + y = 0)$$

και

$$(\exists y)(\forall x)(x + y = 0).$$

Η πρώτη πρόταση είναι αληθής, ενώ η δεύτερη είναι ψευδής. Συνεπώς, η "σειρά" των quantifiers έχει σημασία.

Η άρνηση απλών προτάσεων με quantifiers κατασκευάζεται ως εξής:

$$\begin{aligned} \text{not}\, ((\forall x)P) \quad &\equiv \quad (\exists x)\, \text{not}(P), \\ \text{not}\, ((\exists x)P) \quad &\equiv \quad (\forall x)\, \text{not}(P). \end{aligned}$$

Όταν οι προτάσεις είναι πιο πολύπλοκες δουλεύουμε ως εξής:

$$\begin{aligned} & \text{not}\, [(\forall x)(\exists y)(x + y = 0)] \\ \equiv\ & (\exists x)\, \text{not}\, [(\exists y)(x + y = 0)] \\ \equiv\ & (\exists x)\, (\forall y)\, \text{not}(x + y = 0) \\ \equiv\ & (\exists x)\, (\forall y)\, (x + y \neq 0). \end{aligned}$$

Η τελευταία πρόταση ισχυρίζεται ότι υπάρχει x τέτοιο ώστε για κάθε y το άθροισμα $x + y$ είναι διαφορετικό από το μηδέν. Η πρόταση αυτή είναι ψευδής.