

# Κεφάλαιο 10

## Κυρτή βελτιστοποίηση με περιορισμούς ανισοτήτων

Στο Κεφάλαιο αυτό, μελετούμε προβλήματα κυρτής βελτιστοποίησης στη γενική τους μορφή, δηλαδή, προβλήματα ελαχιστοποίησης κυρτών συναρτήσεων με περιορισμούς ανισοτήτων κυρτών συναρτήσεων και περιορισμούς ισοτήτων αφφινικών συναρτήσεων.

Υπάρχουν αρκετές προσεγγίσεις για την επίλυση αυτών των σημαντικών προβλημάτων.

Θα εστιάσουμε στη μέθοδο εσωτερικού σημείου, αναφέροντας ότι είναι από τις πιο διαδεδομένες μεθόδους επίλυσης προβλημάτων κυρτής βελτιστοποίησης.

### 10.1 Βελτιστοποίηση με περιορισμούς ανισοτήτων

Έστω το πρόβλημα κυρτής βελτιστοποίησης

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & && \mathbf{Ax} = \mathbf{b}, \end{aligned} \tag{10.1}$$

όπου  $f_i : \text{dom } f_i \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , για  $i = 0, \dots, m$ , είναι κυρτές, διπλά διαφορίσιμες συναρτήσεις (άρα, τα  $\text{dom } f_i$ , για  $i = 0, \dots, m$ , είναι ανοιχτά, κυρτά σύνολα),  $\mathbf{A} \in \mathbb{R}^{p \times n}$ , με  $\text{rank}(\mathbf{A}) = p$ , και  $\mathbf{b} \in \mathbb{R}^p$ .

Το σύνολο για το οποίο ορίζεται το πρόβλημα είναι το  $\mathbb{D} := \bigcap_{i=0}^m \text{dom } f_i$ , ενώ το εφικτό σύνολο είναι το

$$\mathbb{X} := \{\mathbf{x} \in \mathbb{D} \mid f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m, \quad \mathbf{Ax} = \mathbf{b}\}. \tag{10.2}$$

Υποθέτουμε ότι το πρόβλημα έχει λύση, έστω  $\mathbf{x}_*$ , και ορίζουμε  $p_* := f_0(\mathbf{x}_*)$ .

Υποθέτουμε ότι το πρόβλημα (10.1) είναι αυστηρά εφικτό. Δηλαδή, υπάρχει  $\mathbf{x} \in \mathbb{D}$  τέτοιο ώστε  $\mathbf{Ax} = \mathbf{b}$  και  $f_i(\mathbf{x}) < 0$ , για  $i = 1, \dots, m$ .

Σε αυτή την περίπτωση, οι συνθήκες KKT είναι αναγκαίες και ικανές συνθήκες βελτιστότητας.

Αν ορίσουμε  $\boldsymbol{\lambda}_* := [\lambda_{*,1} \cdots \lambda_{*,m}]^T$ , τότε οι συνθήκες KKT εκφράζονται ως εξής:

$$\begin{aligned} \nabla f_0(\mathbf{x}_*) + \sum_{i=1}^m \lambda_{*,i} \nabla f_i(\mathbf{x}_*) + \mathbf{A}^T \mathbf{v}_* &= \mathbf{0} \\ \mathbf{Ax}_* &= \mathbf{b} \\ f_i(\mathbf{x}_*) &\leq 0, \quad i = 1, \dots, m \\ \boldsymbol{\lambda}_* &\geq \mathbf{0} \\ \lambda_{*,i} f_i(\mathbf{x}_*) &= 0, \quad i = 1, \dots, m. \end{aligned} \tag{10.3}$$

## 10.2 Μέθοδοι εσωτερικού σημείου

Οι μέθοδοι εσωτερικού σημείου (interior point methods) επιλύουν το πρόβλημα (10.1) λύνοντας μία ακολουθία προβλημάτων με περιορισμούς αφφινικών ισοτήτων.

Τα προβλήματα αυτά είναι “προσεγγίσεις” του προβλήματος (10.1) και οι λύσεις τους είναι αυστηρά εφικτές για το πρόβλημα (10.1).

Στη συνέχεια, περιγράφουμε τη διαδικασία.

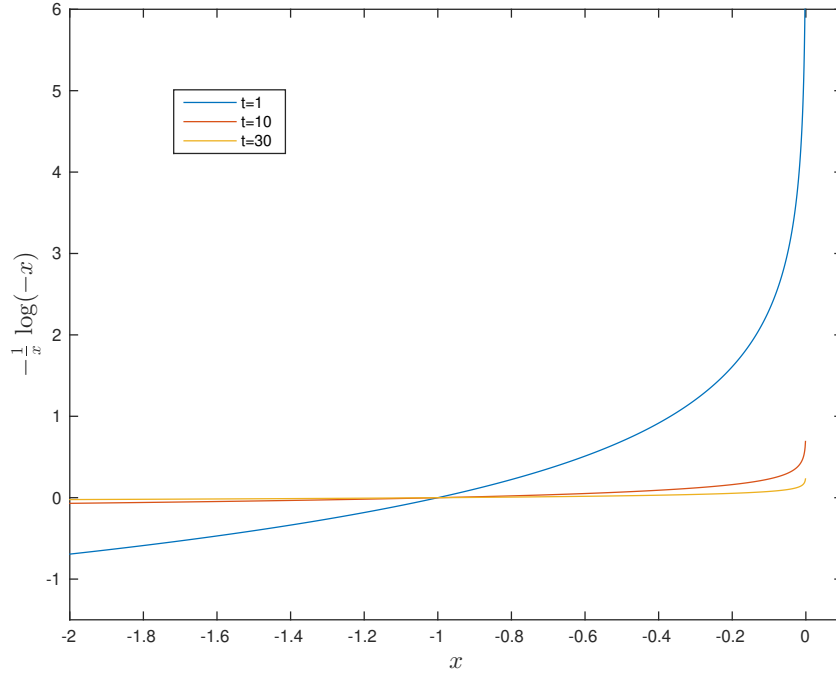
## 10.3 Συνάρτηση λογαριθμικού φράγματος

Το πρόβλημα (10.1) μπορεί να εκφραστεί ως

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) + \sum_{i=1}^m I_-(f_i(\mathbf{x})) \\ \text{subject to} \quad & \mathbf{Ax} = \mathbf{b}, \end{aligned} \tag{10.4}$$

με

$$I_-(u) := \begin{cases} 0, & u \leq 0, \\ \infty, & u > 0. \end{cases} \tag{10.5}$$



Σχήμα 10.1: Προσεγγίσεις της συνάρτησης  $I_-(u)$ , για  $t = 1, 10, 30$ .

Η συνάρτηση κόστους του προβλήματος (10.4), γενικά, είναι μη διαφορίσιμη, συνεπώς, δεν μπορούν να χρησιμοποιηθούν οι μέθοδοι βαθμίδας ή Newton για την επίλυσή του.

Μία προσέγγιση της συνάρτησης  $I_-(u)$  είναι η  $\hat{I}_-(u)$ , η οποία ορίζεται ως

$$\hat{I}_-(u) := -\frac{1}{t} \log(-u), \quad (10.6)$$

με  $\mathbf{dom} \hat{I}_- = -\mathbb{R}_{++}$  και  $t > 0$ .

Η παράμετρος  $t$  καθορίζει την ποιότητα προσέγγισης (δείτε το Σχήμα 10.1).

Η συνάρτηση  $\hat{I}_-$  είναι κυρτή και κλειστή (τείνει στο άπειρο όταν το  $u$  τείνει στο 0) και θα φανεί εξαιρετικά χρήσιμη στη συνέχεια.

Αρχικά, ορίζουμε τη **συνάρτηση λογαριθμικού φράγματος** (logarithmic barrier function) για το πρόβλημα (10.1)

$$\phi(\mathbf{x}) := -\sum_{i=1}^m \log(-f_i(\mathbf{x})), \quad (10.7)$$

με  $\mathbf{dom} \phi = \{\mathbf{x} \in \bigcap_{i=1}^m \mathbf{dom} f_i \mid f_i(\mathbf{x}) < 0, i = 1, \dots, m\}$ . Όπως θα αποδείξουμε στη συνέχεια, η συνάρτηση  $\phi$  είναι κυρτή. Επιπλέον, είναι διπλά διαφορίσιμη.

Στη συνέχεια, ορίζουμε το πρόβλημα βελτιστοποίησης

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) + \frac{1}{t} \phi(\mathbf{x}) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b}, \end{aligned} \quad (10.8)$$

για  $\mathbf{x} \in \mathbf{dom} f_0 \cap \mathbf{dom} \phi$ .<sup>1</sup>

Το πρόβλημα (10.8) είναι κυρτό και ο αλγόριθμος Newton για προβλήματα με περιορισμούς αφφινικών ισοτήτων μπορεί να χρησιμοποιηθεί για την επίλυσή του.

Προφανώς, και άλλοι αλγόριθμοι μπορούν να χρησιμοποιηθούν για την επίλυση του προβλήματος (10.8).

Εμείς, θα εστιάσουμε στον αλγόριθμο Newton.

Όπως είδαμε, και η projected gradient (με περιορισμούς αφφινικών ισοτήτων) μπορεί να χρησιμοποιηθεί.

Το πρόβλημα (10.8) ή το ισοδύναμο πρόβλημα

$$\begin{aligned} & \text{minimize} && t f_0(\mathbf{x}) + \phi(\mathbf{x}) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b} \end{aligned} \quad (10.9)$$

είναι προσεγγίσεις του προβλήματος (10.1), “με την προσέγγιση να βελτιώνεται όσο αυξάνεται το  $t$ .”

Η πρόσθεση της συνάρτησης  $\frac{1}{t} \phi(t)$  στη συνάρτηση κόστους  $f_0(\mathbf{x})$ , διαισθητικά, “υψώνει ένα φράγμα” (barrier) στο σύνορο του εφικτού συνόλου  $\mathbb{X}$ , ουσιαστικά, “παγιδεύοντας” τη λύση του προβλήματος (10.9) στο εσωτερικό του  $\mathbb{X}$  (αλλά επιτρέποντας να φθάσει “πολύ κοντά” στο σύνορό του, για “μεγάλο”  $t$ ).

**Παράδειγμα 10.3.1.** Έστω το απλό πρόβλημα βελτιστοποίησης με περιορισμούς ανισοτήτων

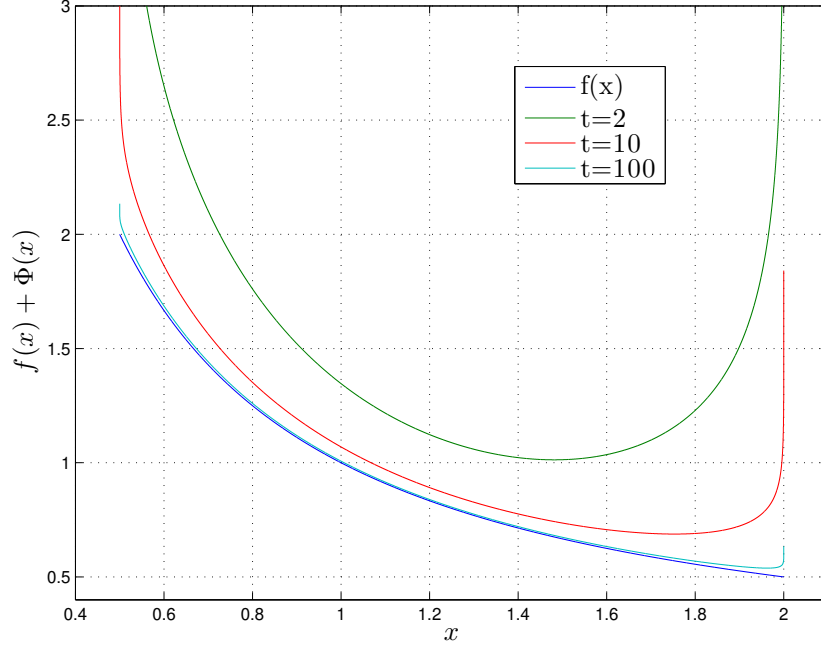
$$\begin{aligned} & \text{minimize} && f_0(x) = \frac{1}{x} \\ & \text{subject to} && 0.5 \leq x \leq 2. \end{aligned} \quad (10.10)$$

Η συνάρτηση  $\phi(x)$  στην περίπτωση αυτή είναι η

$$\phi(x) = -\log(-(x-2)) - \log(-(0.5-x)). \quad (10.11)$$

---

<sup>1</sup>Παρατηρήστε ότι  $\mathbf{dom} f_0 \cap \mathbf{dom} \phi \subset \mathbb{X}$ . Για την ακρίβεια, το σύνολο  $\mathbb{X}$  είναι η κλειστή θήκη του  $\mathbf{dom} f_0 \cap \mathbf{dom} \phi$ .



Σχήμα 10.2: Συναρτήσεις  $f(x) + \frac{1}{t} \phi(x)$ , για  $t = 1, 10, 100$ .

Στο Σχήμα 10.2, σχεδιάζουμε τη συνάρτηση  $f(x) + \frac{1}{t} \phi(x)$ , για  $t = 2, 10, 100$ .

Παρατηρούμε ότι, όταν αυξάνεται το  $t$ , η συνάρτηση  $f(x) + \frac{1}{t} \phi(x)$  πλησιάζει την  $f(\mathbf{x})$  εκτός από τα άκρα του διαστήματος  $[0.5, 2]$  όπου αυξάνεται απότομα, σχηματίζοντας ένα “φράγμα.”  $\square$

Η πρώτη και η δεύτερη παράγωγος της  $\phi(\mathbf{x})$  έχουν ως εξής:

$$\begin{aligned} \nabla \phi(\mathbf{x}) &= \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}), \\ \nabla^2 \phi(\mathbf{x}) &= \sum_{i=1}^m \frac{1}{f_i^2(\mathbf{x})} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^T - \sum_{i=1}^m \frac{1}{f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}). \end{aligned} \quad (10.12)$$

Από τη μορφή της  $\nabla^2 \phi(\mathbf{x})$  και από το γεγονός ότι  $f_i(\mathbf{x}) < 0$  για  $\mathbf{x} \in \mathbf{dom} \phi$ , διαπιστώνουμε εύκολα ότι η  $\phi(\mathbf{x})$  είναι κυρτή συνάρτηση (διότι η Hessian είναι μη-αρνητικά ορισμένη).

## 10.4 Κεντρικό μονοπάτι

Υποθέτουμε ότι, για  $t > 0$ , υπάρχει λύση του προβλήματος (10.9), έστω  $\mathbf{x}_*(t)$  (συνεπώς,  $f_0(\mathbf{x}_*(t)) > -\infty$ ).

Η συνάρτηση  $\mathbf{x}_*(t)$  καλείται **κεντρικό μονοπάτι** (central path), ενώ κάθε σημείο του κεντρικού μονοπατιού καλείται **κεντρικό σημείο** (central point).

Αν το  $\mathbf{x}_*(t)$  είναι κεντρικό σημείο, τότε υπάρχει  $\hat{\mathbf{v}} \in \mathbb{R}^p$  τέτοιο ώστε να ικανοποιούνται οι σχέσεις KKT για το πρόβλημα (10.9), δηλαδή,

$$t \nabla f_0(\mathbf{x}_*(t)) + \nabla \phi(\mathbf{x}_*(t)) + \mathbf{A}^T \hat{\mathbf{v}} = \mathbf{0}, \quad \mathbf{A} \mathbf{x}_*(t) = \mathbf{b}, \quad (10.13)$$

ή, ισοδύναμα,

$$t \nabla f_0(\mathbf{x}_*(t)) + \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x}_*(t))} \nabla f_i(\mathbf{x}_*(t)) + \mathbf{A}^T \hat{\mathbf{v}} = \mathbf{0}, \quad \mathbf{A} \mathbf{x}_*(t) = \mathbf{b}. \quad (10.14)$$

Επιπλέον,  $\mathbf{x}_*(t) \in \text{dom } f_0 \cap \text{dom } \phi$ , συνεπώς,

$$f_i(\mathbf{x}_*(t)) < 0, \quad \text{για } i = 1, \dots, m. \quad (10.15)$$

#### 10.4.1 Εφικτά δυϊκά σημεία από τα σημεία του κεντρικού μονοπατιού

Από την (10.14), μπορούμε να εξάγουμε το εξής σημαντικό συμπέρασμα για το κεντρικό μονοπάτι.

Κάθε κεντρικό σημείο  $\mathbf{x}_*(t)$  ορίζει ένα εφικτό δυϊκό σημείο για το πρόβλημα (10.1) και, συνεπώς, ένα κάτω φράγμα για το  $p_*$ .

Πιο συγκεκριμένα, έστω

$$\lambda_{*,i}(t) := -\frac{1}{t f_i(\mathbf{x}_*(t))}, \quad \text{για } i = 1, \dots, m, \quad \mathbf{v}_*(t) = \frac{\hat{\mathbf{v}}}{t}. \quad (10.16)$$

Μπορεί να αποδειχθεί ότι το σημείο  $(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t))$ , με  $\boldsymbol{\lambda}_*(t) := [\lambda_{*,1}(t) \cdots \lambda_{*,m}(t)]^T$ , είναι εφικτό δυϊκό σημείο για το πρόβλημα (10.1).<sup>2</sup>

Η απόδειξη έχει ως εξής.

Εξ ορισμού, έχουμε ότι  $\boldsymbol{\lambda}_*(t) > \mathbf{0}$ , διότι  $t > 0$  και  $f_i(\mathbf{x}_*(t)) < 0$ , για  $i = 1, \dots, m$ .

Η σχέση (10.14) μπορεί να εκφραστεί ως

$$\nabla f_0(\mathbf{x}_*(t)) + \sum_{i=1}^m \lambda_{*,i}(t) \nabla f_i(\mathbf{x}_*(t)) + \mathbf{A}^T \mathbf{v}_*(t) = \mathbf{0}. \quad (10.17)$$

---

<sup>2</sup>Υπενθυμίζουμε ότι ένα σημείο  $(\boldsymbol{\lambda}, \mathbf{v})$  είναι δυϊκό εφικτό αν  $\boldsymbol{\lambda} \geq \mathbf{0}$  και  $g(\boldsymbol{\lambda}, \mathbf{v}) > -\infty$ .

Η συνάρτηση Lagrange για το πρόβλημα (10.1) έχει ως εξής

$$L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = f_0(\mathbf{x}) + \sum_{i=1}^n \lambda_i f_i(\mathbf{x}) + \mathbf{v}^T (\mathbf{A}\mathbf{x} - \mathbf{b}). \quad (10.18)$$

Έστω  $\boldsymbol{\lambda} = \boldsymbol{\lambda}_*(t)$  και  $\mathbf{v} = \mathbf{v}_*(t)$ . Τότε, γνωρίζουμε ότι

$$g(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t)) = \inf_{\mathbf{x} \in \mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}_*(t), \mathbf{v}_*(t)). \quad (10.19)$$

Η συνάρτηση  $L(\mathbf{x}, \boldsymbol{\lambda}_*(t), \mathbf{v}_*(t))$  είναι κυρτή συνάρτηση του  $\mathbf{x}$  (γιατί;).

Συνεπώς, μία αναγκαία και ικανή συνθήκη για την ελαχιστοποίησή της πάνω στο ανοιχτό σύνολο  $\mathbb{D}$  είναι η

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}_*(t), \mathbf{v}_*(t)) = \mathbf{0}. \quad (10.20)$$

Διαπιστώνουμε εύκολα ότι, εξαιτίας της (10.17), το  $\mathbf{x}_*(t)$  ικανοποιεί την (10.20), συνεπώς, ελαχιστοποιεί την  $L(\mathbf{x}, \boldsymbol{\lambda}_*(t), \mathbf{v}_*(t))$ .

Η δυϊκή συνάρτηση στη θέση  $(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t))$  είναι η

$$\begin{aligned} g(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t)) &= L(\mathbf{x}_*(t), \boldsymbol{\lambda}_*(t), \mathbf{v}_*(t)) \\ &= f_0(\mathbf{x}_*(t)) + \sum_{i=1}^m \lambda_{*,i}(t) f_i(\mathbf{x}_*(t)) + \mathbf{v}_*(t)^T (\mathbf{A}\mathbf{x}_*(t) - \mathbf{b}) \\ &\stackrel{(10.16)}{=} f_0(\mathbf{x}_*(t)) - \frac{m}{t}. \end{aligned} \quad (10.21)$$

Εξαιτίας του ότι  $f_0(\mathbf{x}_*(t)) > -\infty$ , έχουμε ότι  $g(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t)) > -\infty$ .

Άρα, το ζεύγος  $(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t))$  είναι εφικτό δυϊκό σημείο για το πρόβλημα (10.1).

Συνεπώς,

$$g(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t)) \leq p_*. \quad (10.22)$$

Συνδυάζοντας τις σχέσεις (10.21) και (10.22), λαμβάνουμε την εξαιρετικά σημαντική ανισότητα

$$f_0(\mathbf{x}_*(t)) - p_* \leq \frac{m}{t}. \quad (10.23)$$

Δηλαδή,

1. για δεδομένο  $t$ , το σημείο  $\mathbf{x}_*(t)$  είναι το πολύ  $\frac{m}{t}$ -υποβέλτιστο για το πρόβλημα (10.1),
2. για  $t \rightarrow \infty$ , το  $\mathbf{x}_*(t)$  συγκλίνει σε ένα βέλτιστο σημείο για το πρόβλημα (10.1).

### 10.4.2 Ερμηνεία μέσω ΚΚΤ

Είδαμε ότι οι σχέσεις

$$\begin{aligned} \mathbf{Ax} &= \mathbf{b}, \quad f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ \boldsymbol{\lambda} &\geq \mathbf{0} \\ \nabla f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} &= \mathbf{0} \\ \lambda_i f_i(\mathbf{x}) &= -\frac{1}{t}, \quad i = 1, \dots, m \end{aligned} \tag{10.24}$$

ικανοποιούνται από τα  $\mathbf{x}_*(t)$ ,  $\boldsymbol{\lambda}_*(t)$ , και  $\mathbf{v}_*(t)$ .

Οι σχέσεις (10.24) είναι οι συνθήκες ΚΚΤ για το πρόβλημα (10.1), με μόνη διαφορά το ότι αντί για μηδέν στο δεξιό μέλος της τέταρτης γραμμής υπάρχει ο όρος  $-\frac{1}{t}$ .

Συνεπώς, για “μεγάλο”  $t$ , τα  $\mathbf{x}_*(t)$ ,  $\boldsymbol{\lambda}_*(t)$ , και  $\mathbf{v}_*(t)$  “σχεδόν” ικανοποιούν τις συνθήκες ΚΚΤ για το πρόβλημα (10.1).

## 10.5 Η μέθοδος φράγματος

Είδαμε ότι το σημείο  $\mathbf{x}_*(t)$  είναι το πολύ  $\frac{m}{t}$ -υποβέλτιστο για το πρόβλημα (10.1).

Μία προσέγγιση για την επίλυση του (10.1), με ανοχή  $\epsilon$ , είναι η λύση του προβλήματος

$$\begin{aligned} \text{minimize} \quad & \frac{m}{\epsilon} f_0(\mathbf{x}) + \phi(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{Ax} = \mathbf{b}. \end{aligned} \tag{10.25}$$

Γενικά, το πρόβλημα αυτό είναι δύσκολο να λυθεί για “μικρό”  $\epsilon$ . Συνήθως, ακολουθείται η διαδικασία η οποία περιγράφεται στη συνέχεια.

## 10.6 Η μέθοδος φράγματος

Μία αποδοτική προσέγγιση για την επίλυση του προβλήματος (10.25) αποτελεί η επίλυση ακολουθίας προβλημάτων της μορφής (10.9), με αυξανόμενο  $t$ .

Η λύση του προβλήματος για ένα συγκεκριμένο  $t$  αποτελεί το σημείο εκκίνησης για το επόμενο πρόβλημα (με μεγαλύτερο  $t$ ).



Η μέθοδος αυτή, η οποία καλείται **μέθοδος φράγματος** (barrier method), περιγράφεται στον Πίνακα 10.6.

Αρχικά, είχε προταθεί από τους Fiacco και McCormick τη δεκαετία του 1960.

Όμως, κατά την επόμενη δεκαετία, δεν γνώρισε ευρεία αποδοχή κυρίως διότι υπήρχαν επιχειρήματα κατά της αριθμητικής ακρίβειας της λύσης της μεθόδου.

Για παράδειγμα, για μεγάλο  $t$ , ο αριθμός συνθήκης της δεύτερης παραγώγου της συνάρτησης  $tf_0 + \phi$  είναι πολύ μεγάλος, γεγονός το οποίο αναμενόταν να οδηγήσει σε δυσκολίες κατά την αριθμητική υλοποίηση της μεθόδου.

Η μέθοδος επανήλθε στο προσκήνιο κατά τη δεκαετία του 1980, όταν συνδέθηκε με τον αλγόριθμο του Karmarkar για την επίλυση προβλημάτων γραμμικού προγραμματισμού.

Τα τελευταία δεκαπέντε χρόνια, αποτελεί μία από τις πιο διαδεδομένες προσεγγίσεις για τη λύση προβλημάτων κυρτής βελτιστοποίησης.

---

$\mathbf{x} \in \text{dom } f_0 \cap \text{dom } \phi$ ,  $\mathbf{Ax} = \mathbf{b}$ ,  $t > 0$ ,  $\mu > 1$ , tolerance  $\epsilon > 0$ .

repeat

1. Compute  $\mathbf{x}_*(t)$ , by minimizing  $tf_0 + \phi$  subject to  $\mathbf{Ax} = \mathbf{b}$ , starting at  $\mathbf{x}$ .
  2.  $\mathbf{x} := \mathbf{x}_*(t)$ .
  3. quit if  $\frac{m}{t} < \epsilon$ .
  3.  $t := \mu t$ .
- 

Πίνακας 10.1: Μέθοδος φράγματος για προβλήματα κυρτής βελτιστοποίησης (εκκίνηση από εφικτό σημείο).

### 10.6.1 Ακρίβεια του κεντρικού βήματος

Δείτε στο βιβλίο, σελίδα 570.

### 10.6.2 Επιλογή της παραμέτρου $\mu$

Δείτε στο βιβλίο, σελίδα 570.

### 10.6.3 Επιλογή της αρχικής τιμής της παραμέτρου $t$

Δείτε στο βιβλίο, σελίδα 570.

## 10.7 Παραδείγματα

Παράδειγμα 10.7.1. Γραμμικός προγραμματισμός

Θα το μελετήσετε σε βάθος στην 5η άσκηση.

## 10.8 Ανάλυση σύγκλισης

Δείτε στο βιβλίο, σελίδα 577.

## 10.9 Βήμα Newton για τροποποιημένες ΚΚΤ

Όχι.

## 10.10 Εφικτότητα και Φάση I

Μία προσέγγιση για την εύρεση του αρχικού εφικτού σημείου για το πρόβλημα (10.1) είναι η λύση του προβλήματος εφικτότητας

$$\begin{aligned} & \text{minimize} && s \\ & \text{subject to} && f_i(\mathbf{x}) \leq s, \quad i = 1, \dots, m \\ & && \mathbf{Ax} = \mathbf{b}, \end{aligned} \tag{10.26}$$

με μεταβλητές βελτιστοποίησης  $\mathbf{x} \in \mathbb{R}^n$ ,  $s \in \mathbb{R}$  (θυμίζουμε ότι  $\mathbf{A} \in \mathbb{R}^{p \times n}$  και  $\mathbf{b} \in \mathbb{R}^p$ ).

Αν υπάρχει  $\mathbf{x}$  τέτοιο ώστε  $\mathbf{Ax} = \mathbf{b}$ , έστω  $\bar{\mathbf{x}}$ , τότε το πρόβλημα (10.26) είναι πάντα αυστηρά εφικτό, διότι μπορούμε να θέσουμε ως αρχικό σημείο το  $(\bar{\mathbf{x}}, \bar{s})$  με

$$\bar{s} \geq \max\{f_1(\bar{\mathbf{x}}), \dots, f_m(\bar{\mathbf{x}})\}. \tag{10.27}$$

Αν  $p < n$ , τότε το σύστημα  $\mathbf{Ax} = \mathbf{b}$  έχει άπειρες λύσεις (θυμηθείτε ότι έχουμε υποθέσει ότι οι γραμμές του  $\mathbf{A}$  είναι γραμμικά ανεξάρτητες). Μία λύση δίνεται από τη σχέση  $\bar{\mathbf{x}} = \mathbf{A}^\# \mathbf{b}$ , όπου  $\mathbf{A}^\#$  είναι ο ψευδοαντίστροφος του  $\mathbf{A}$ .

Έστω  $\bar{\mathbf{x}}_*$  το βέλτιστο σημείο και  $\bar{p}_*$  η λύση του προβλήματος (10.26). Τότε

1. αν  $\bar{p}_* > 0$ , τότε το πρόβλημα (10.1) είναι *ανέφικτο*.
2. αν  $\bar{p}_* < 0$ , τότε το πρόβλημα (10.1) είναι εφικτό και ένα εφικτό σημείο για το πρόβλημα (10.1) είναι το  $\bar{\mathbf{x}}_*$ .
3. αν  $\bar{p}_* = 0$ , τότε το πρόβλημα δεν είναι αυστηρά εφικτό.

Στην πράξη, αυτό το οποίο είναι δυνατό να συμβεί είναι το να έχουμε  $|\bar{p}_*| < \epsilon$ , για ένα πολύ μικρό  $\epsilon > 0$ .

Τότε, οι ανισότητες  $f_i(\mathbf{x}) \leq -\epsilon$ , για  $i = 1, \dots, m$ , δεν είναι εφικτές, ενώ οι ανισότητες  $f_i(\mathbf{x}) \leq \epsilon$ , για  $i = 1, \dots, m$ , είναι εφικτές.

Θα πρέπει να σημειώσουμε ότι δεν είναι απαραίτητο να λύσουμε το πρόβλημα (10.26).

Αρκεί να βρούμε ένα  $\mathbf{x}$  το οποίο να οδηγεί σε  $s < 0$ .

## 10.11 Φάση Ι μέσω αλγορίθμου Newton με εκκίνηση από μη εφικτό σημείο

## 10.12 Άλλες συναρτήσεις barrier

Εκτός της

$$\phi(\mathbf{x}) = - \sum_{j=1}^m \log(-f_j(\mathbf{x})),$$

μπορούν να χρησιμοποιηθούν και άλλες συναρτήσεις, όπως οι εξής:

1. Power-function barrier:

$$\phi(\mathbf{x}) = \sum_{j=1}^m \frac{1}{(-f_j(\mathbf{x}))^p}, \quad p \geq 1.$$

2. Exponential barrier:

$$\phi(\mathbf{x}) = \sum_{j=1}^m \exp\left(\frac{1}{-f_j(\mathbf{x})}\right).$$

Οι παραπάνω συναρτήσεις οδηγούν σε interior point μεθόδους της μορφής ( $t \rightarrow \infty$ ):

$$\min_{\mathbf{x}} f_0(\mathbf{x}) + \frac{1}{t} \phi(\mathbf{x}), \quad \text{s.t. } \mathbf{Ax} = \mathbf{b}.$$

### 10.13 Συναρτήσεις Penalty

Εστω η συνάρτηση

$$(x)_+ = \max\{0, x\} := \begin{cases} 0, & x \leq 0. \\ x, & x > 0. \end{cases} \quad (10.28)$$

1. Συνάρτηση τετραγωνικού penalty:

$$P(\mathbf{x}) = \sum_{j=1}^m (f_j(\mathbf{x}))_+^2.$$

2. Μη-διαφορίσιμο penalty:

$$P(\mathbf{x}) = \sum_{j=1}^m (f_j(\mathbf{x}))_+.$$

Οι παραπάνω συναρτήσεις οδηγούν σε exterior point μεθόδους της μορφής ( $t \rightarrow \infty$ ):

$$\min_{\mathbf{x}} f_0(\mathbf{x}) + t P(\mathbf{x}), \quad \text{s.t. } \mathbf{Ax} = \mathbf{b}.$$

### 10.14 Πρωτογενής – Δυϊκή μέθοδος εσωτερικού σημείου

### 10.15 Πρωτογενής – δυϊκή κατεύθυνση κίνησης

Έστω

$$\mathbf{f}(\mathbf{x}) := \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}, \quad D\mathbf{f}(\mathbf{x}) := \begin{bmatrix} \nabla f_1(\mathbf{x})^T \\ \vdots \\ \nabla f_m(\mathbf{x})^T \end{bmatrix}. \quad (10.29)$$

Ορίζουμε το διάνυσμα  $r_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v})$  ως εξής:

$$r_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) := \begin{bmatrix} \nabla f_0(\mathbf{x}) + D\mathbf{f}(\mathbf{x})^T \boldsymbol{\lambda} + \mathbf{A}^T \mathbf{v} \\ -\text{diag}(\boldsymbol{\lambda})\mathbf{f}(\mathbf{x}) - \frac{1}{t} \mathbf{1} \\ \mathbf{Ax} - \mathbf{b} \end{bmatrix}. \quad (10.30)$$

Αν τα  $\mathbf{x} \in \mathbf{dom} f_0 \cap \mathbf{dom} \phi$ ,  $\boldsymbol{\lambda} > \mathbf{0}$ , και  $\mathbf{v}$  ικανοποιούν την εξίσωση  $r_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = \mathbf{0}$ , τότε  $\mathbf{x} = \mathbf{x}_*(t)$ ,  $\boldsymbol{\lambda} = \boldsymbol{\lambda}_*(t)$ , και  $\mathbf{v} = \mathbf{v}_*(t)$ .

Ειδικότερα, το  $\mathbf{x}_*(t)$  είναι εφικτό πρωτογενές σημείο και το  $(\boldsymbol{\lambda}_*(t), \mathbf{v}_*(t))$  είναι εφικτό δυϊκό σημείο, με χάσμα δυϊκότητας  $\frac{m}{t}$ .

Έστω ότι ευρισκόμαστε στο σημείο  $\mathbf{y} = (\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v})$ , με  $\mathbf{x} \in \mathbf{dom} f_0 \cap \mathbf{dom} \phi$  και  $\boldsymbol{\lambda} > \mathbf{0}$ .

Το διάνυσμα

$$r_{t,d}(\mathbf{y}) := \nabla f_0(\mathbf{x}) + D\mathbf{f}(\mathbf{x})^T \boldsymbol{\lambda} + \mathbf{A}^T \mathbf{v} \quad (10.31)$$

καλείται **δυϊκό υπόλοιπο** (dual residual), το διάνυσμα

$$r_{t,p}(\mathbf{y}) := \mathbf{A}\mathbf{x} - \mathbf{b} \quad (10.32)$$

καλείται **πρωτογενές υπόλοιπο** (primal residual), ενώ το διάνυσμα

$$r_{t,c}(\mathbf{y}) := -\mathbf{diag}(\boldsymbol{\lambda})\mathbf{f}(\mathbf{x}) - \frac{1}{t} \mathbf{1}, \quad (10.33)$$

όπου  $\mathbf{1}$  είναι το  $m$ -διάστατο διάνυσμα με στοιχεία μονάδες, καλείται **υπόλοιπο κεντρικότητας** (centrality residual).

Η πρώτη τάξης προσέγγιση του  $r_t$ , στο σημείο  $\mathbf{y}$ , έχει ως εξής

$$r_t(\mathbf{y} + \mathbf{z}) \approx r_{t,\mathbf{y}}(\mathbf{z}) = r_t(\mathbf{y}) + Dr_t(\mathbf{y})\mathbf{z}. \quad (10.34)$$

Ορίζουμε ως βήμα Newton  $\Delta\mathbf{y} = (\Delta\mathbf{x}, \Delta\boldsymbol{\lambda}, \Delta\mathbf{v})$ , στο σημείο  $\mathbf{y}$ , το διάνυσμα  $\mathbf{z}$  για το οποίο μηδενίζεται η  $r_{t,\mathbf{y}}(\mathbf{z})$ , δηλαδή

$$Dr_t(\mathbf{y})\Delta\mathbf{y} = -r_t(\mathbf{y}). \quad (10.35)$$

Υπολογίζοντας την παράγωγο  $Dr_t$ , στο σημείο  $\mathbf{y}$ , λαμβάνουμε

$$\begin{bmatrix} \nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(\mathbf{x}) & D\mathbf{f}(\mathbf{x})^T & \mathbf{A}^T \\ -\mathbf{diag}(\boldsymbol{\lambda})D\mathbf{f}(\mathbf{x}) & -\mathbf{diag}(\mathbf{f}(\mathbf{x})) & \mathbf{O} \\ \mathbf{A} & \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda} \\ \Delta\mathbf{v} \end{bmatrix} = - \begin{bmatrix} r_{t,d}(\mathbf{y}) \\ r_{t,c}(\mathbf{y}) \\ r_{t,p}(\mathbf{y}) \end{bmatrix}. \quad (10.36)$$

Ορίζουμε ως **πρωτογενή – δυϊκή κατεύθυνση κίνησης** (primal-dual search direction),  $\Delta\mathbf{y}_{pd} := (\Delta\mathbf{x}_{pd}, \Delta\boldsymbol{\lambda}_{pd}, \Delta\mathbf{v}_{pd})$ , το βήμα Newton  $\Delta\mathbf{y}$ .

Οι μεταβλητές  $\mathbf{x}_k$  και  $(\boldsymbol{\lambda}_k, \mathbf{v}_k)$  του πρωτογενούς – δυϊκού αλγορίθμου εσωτερικού σημείου δεν είναι αναγκαστικά εφικτά σημεία του προβλήματος (10.1), παρά μόνο οριακά (δηλαδή, στη σύγκλιση του αλγορίθμου).

Αυτό σημαίνει ότι δεν έχουμε κάποιο προφανή τρόπο για να υπολογίζουμε το χάσμα δυϊκότητας κατά τη διάρκεια εκτέλεσης του αλγορίθμου.

Για  $\mathbf{x} \in \text{dom } f_0 \cap \text{dom } \phi$  και  $\boldsymbol{\lambda} > \mathbf{0}$ , ορίζουμε το **αναπληρωματικό χάσμα δυϊκότητας** (surrogate duality gap) ως

$$\hat{\eta}(\mathbf{x}, \boldsymbol{\lambda}) := -\mathbf{f}(\mathbf{x})^T \boldsymbol{\lambda}. \quad (10.37)$$

Η ποσότητα  $\hat{\eta}(\mathbf{x}, \boldsymbol{\lambda})$  θα ήταν ίση με το χάσμα δυϊκότητας ( $= -\frac{m}{t}$ ) αν τα  $\mathbf{x}$  και  $\boldsymbol{\lambda}$  ήταν εφικτά σημεία για το πρόβλημα (10.1).

Η τιμή του  $t$  η οποία αντιστοιχεί σε αναπληρωματικό χάσμα δυϊκότητας  $\hat{\eta}$  ισούται με  $\frac{m}{\hat{\eta}}$ .

### 10.15.1 Αναζήτηση γραμμής

Εστω

$$\mathbf{x}_+ := \mathbf{x} + s\Delta\mathbf{x}_{\text{pd}}, \quad \boldsymbol{\lambda}_+ := \boldsymbol{\lambda} + s\Delta\boldsymbol{\lambda}_{\text{pd}}, \quad \mathbf{v}_+ := \mathbf{v} + s\Delta\mathbf{v}_{\text{pd}}. \quad (10.38)$$

Η αναζήτηση γραμμής θα πρέπει να εγγυηθεί ότι  $\boldsymbol{\lambda}_+ > \mathbf{0}$ ,  $\mathbf{f}(\mathbf{x}_+) < \mathbf{0}$  και η συνάρτηση  $\|r(\cdot)\|_2$  μειώνεται αρκετά (δες σελίδα 613 Boyd & Vandenberghe).

1. Υπολόγισε την ποσότητα  $s^{\max} := \sup\{s \in [0, 1] \mid \boldsymbol{\lambda} + s\Delta\boldsymbol{\lambda} > \mathbf{0}\}$ .
2. Οπισθοχώρησε κατά  $\beta \in (0, 1)$  μέχρι  $\mathbf{f}(\mathbf{x}_+) < \mathbf{0}$ .
3. Οπισθοχώρησε κατά  $\beta \in (0, 1)$  μέχρι

$$\|r_t(\mathbf{x}_+, \boldsymbol{\lambda}_+, \mathbf{v}_+)\|_2^2 \leq (1 - \alpha s) \|r_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v})\|_2^2 \quad (10.39)$$

για  $\alpha \in (0, 0.5)$ .

Προσπαθήστε να αποδείξετε ότι η σχέση (10.39) είναι συμβατή με το backtracking line search όπως το έχετε συναντήσει σε προηγούμενα εδάφια. Τότε, θα καταλάβετε ότι, με τα ίδια επιχειρήματα, μπορεί να αποδειχθεί ότι ο backtracking line search θα τερματίζει πάντα, μετά από πεπερασμένο αριθμό βημάτων.

### 10.15.2 Αλγόριθμος primal-dual για προβλήματα κυρτής βελτιστοποίησης

Ο αλγόριθμος primal-dual παρουσιάζεται στον Πίνακα 10.2.

Παρατηρούμε ότι θα πρέπει να εκκινήσουμε από σημείο το οποίο ικανοποιεί αυστηρά όλους τους περιορισμούς ανισοτήτων.

---

$\mathbf{x} \in \mathbf{dom} f_0 \cap \mathbf{dom} \phi, \boldsymbol{\lambda} > \mathbf{0}, t > 0, \mu > 1, e_{\text{feas}} > 0, e > 0.$

repeat

1.  $t := \frac{m\mu}{\hat{\eta}}.$
2. Compute  $\Delta \mathbf{y}_{\text{pd}}.$
3. perform line search and choose step  $s.$
4.  $\mathbf{y} := \mathbf{y} + s\Delta \mathbf{y}_{\text{pd}}.$

until  $(\|r_{\text{p}}\|_2 < e_{\text{feas}}, \|r_{\text{d}}\|_2 < e_{\text{feas}}, \hat{\eta} < e).$

---

Πίνακας 10.2: Αλγόριθμος primal-dual για προβλήματα κυρτής βελτιστοποίησης.