Αριθμητικές μέθοδοι για προβλήματα βελτιστοποίησης με περιορισμούς / δεσμεύσεις

Κ. Χρυσαφίνος

Η μέθοδος Frank-Wolfe

- Πρόβλημα (ΥΠΠ): Να υπολογιστούν σημεία τοπικού ή / και ολικού ελάχιστου για το πρόβλημα $\min_{x \in S} f(x) = f(\bar{x})$.

- Υποθέτουμε ότι η συνάρτηση $f : \mathbb{R}^n \to \mathbb{R},$ ικανοποιεί $f \in C^1(\mathbb{R}^n)$, ενώ το σύνολο $S \neq \emptyset, S \subset \mathbb{R}^n$ κυρτό και συμπαγές. Επομένως από το Θεώρημα 1 υπάρχει τουλάχιστον ένα σημείο ελαχίστου. Η ιδιότητα του κυρτού υποσυνόλου $S \subset \mathbb{R}^n$ είναι εξαιρετικά σημαντικός στην κατασκευή των αριθμητικών μεθόδων.

- Αναγκαία συνθήκη: $\nabla f(\bar{x})^T(x - \bar{x}) \geq 0, \quad \forall x \in S$.

- Οι υπολογιστικές μέθοδοι, για δοσμένη αρχική συνθήκη $x_0 \in \mathbb{R}^n$, κατασκευάζουν ακολουθία $\{x_k\}_{k=0}^{\infty} \in \mathbb{R}^n$ που (στο όριο - με κατάλληλο τρόπο) προσεγγίζει τη λύση ή λύσεις του προβλήματος (ΥΠΠ), δηλαδή λύσεις που ικανοποιούν την αναγκαία συνθήκη.

- Ο αλγόριθμος Frank-Wolfe

- Με Χρήση του "μετρητή" $\delta_k := \nabla f(x_k)^T(y_k - x_k)$

1. Given $x_1 \in S$. Set $y_0 := x_1$

   For $k = 1,..,$ compute until convergence

2. Compute $y_k \in S$ such that

$$\min_{y \in S} \nabla f(x_k)^T(y - x_k) = \nabla f(x_k)^T(y_k - x_k) := \delta_k$$

3. If $\delta_k = 0$ Stop. Else

4. Compute a_k such that

$$\min_{a \in [0,1]} f(x_k + a(y_k - x_k)) = f(x_k + a_k(y_k - x_k))$$

5. $x_{k+1} = x_k + a_k(y_k - x_k)$

- *Θεώρημα:* Έστω $f \in \mathbb{R}^n \to \mathbb{R}, f \in C^1(\mathbb{R}^n)$. Υποθέτουμε ότι το σύνολο $S \subset \mathbb{R}^n$ είναι κυρτό και συμπαγές. Αν ο αλγόριθμος σταματά για κάποιο κ, τότε $\nabla f(x_k)^T(y - x_k) \geq 0,\ \forall y \in S$ δηλαδή το $\bar{x} := x_k$ ικανοποιεί την αναγκαία συνθήκη. Διαφορετικά, η ακολουθία $\{x_k\}_{k=0,\ldots}$ που παράγει ο αλγόριθμος είναι άπειρη και κάθε όριο $\bar{x}$ συγκλίνουσας υπακολουθίας ικανοποιεί $\nabla f(\bar{x})^T(y - \bar{x}) \geq 0,\ \forall\ y \in S$.

- *Απόδειξη:* Θα δείξουμε επαγωγικά ότι αν $\delta_k \neq 0$, η ακολουθία $\{x_k\}_{k=0,\ldots}$,είναι καλά ορισμένη, (δηλαδή υπάρχει το $x_k \in S$, για κάθε κ) και ότι η $\{f(x_k)\}_{k=0,\ldots}$ είναι φθίνουσα. Έστω $x_1 \in S$. Ορίζουμε $y_0 = x_1$. Παρατηρούμε πως το πρόβλημα του βήματος 2, είναι καλά ορισμένο καθώς η συνάρτηση $y \in S \to \nabla f(x_1)^T(y - x_1) \in \mathbb{R}$ είναι συνεχής (γραμμική ως προς y) ενώ το $S \subset \mathbb{R}^n$ είναι συμπαγές. Επομένως υπάρχει τουλάχιστον ένα σημείο ελαχίστου y_1 στο S , δηλαδή, $\min_{y \in S} \nabla f(x_1)^T(y - x_1) = \nabla f(x_1)^T(y_1 - x_1)$. Παρατηρούμε πως η συνάρτηση $y \in S \to \nabla f(x_1)^T(y - x_1) \in \mathbb{R}$ είναι κυρτή, ορισμένη πάνω σε κυρτό σύνολο, επομένως το ελάχιστο y_1 είναι σημείο (ολικού) ελαχίστου. Εύκολα διαπιστώνουμε πως το βήμα 4 είναι επίσης καλά ορισμένο καθώς πρόκειται για πρόβλημα ελαχιστοποίσης συνεχούς συνάρτησης ορισμένης στο συμπαγές [0,1]. Επίσης επειδή τα $x_1, y_1 \in S$ και $a_1 \in [0,1]$, έχουμε ότι το
$x_2 = x_1 + a_1(y_1 - x_1) = (1 - a_1)x_1 + a_1 y_1 \in S$. ($S \subset \mathbb{R}^n$ κυρτό).

- (Συνέχεια απόδειξης:) Παρατηρούμε επίσης, από τον ορισμό του a_1 πως,
$$\min_{a\in[0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1)) = f(x_2) \leq f(x_1 + 0(y_1 - x_1))$$
δηλαδή, $f(x_2) \leq f(x_1)$.
Επαγωγικά, χρησιμοποιώντας παρόμοια επιχειρήματα, αποδεικνύεται πως η ακολουθία $\{x_k\}_{k=1,...}$,είναι καλά ορισμένη, (δηλαδή υπάρχει το $x_k \in S$, για κάθε κ) και ότι η $\{f(x_k)\}_{k=1,...}$ είναι φθίνουσα. Προφανώς αν ο αλγόριθμος σταματά για κάποιο $\delta_k := \nabla f(x_k)^T(y_k - x_k)$ $= 0$ το $\bar{x} := x_k$ ικανοποιεί την αναγκαία συνθήκη καθώς,
$$0 = \nabla f(x_k)^T(y_k - x_k) := \min_{y\in S} \nabla f(x_k)^T(y - x_k) \leq \nabla f(x_k)^T(x - x_k) \quad \forall x \in S$$
Επομένως, διαφορετικά η ακολουθία είναι άπειρη.

- *(Συνέχεια Απόδειξης):* Επειδή το $S \subset \mathbb{R}^n$είναι συμπαγές και ολόκληρη η ακολουθία $x_k \in S,$ για κάθε κ, έχουμε πως υπάρχει συγκλίνουσα υπακολουθία, δηλαδή υπάρχει $\bar{x} \in S$, $\bar{y} \in S$ ώστε $\lim_{k_n \to \infty} x_{k_n} = \bar{x}$ και $\lim_{k_l \to \infty} y_{k_l} = \bar{y}$.
  Διαλέγοντας του κοινούς δείκτες και επειδή, $f \in C^1(\mathbb{R}^n)$, έχουμε
  $$\lim_{k_{n,l} \to \infty} \nabla f(x_{k_{n,l}})^T (y_{k_{n,l}} - x_{k_{n,l}}) = \nabla f(\bar{x})^T(\bar{y} - \bar{x}) := \delta \leq 0$$
  (Άσκηση: Να δείξετε ότι το $\delta \leq 0$.) Θα δείξουμε ότι $\delta = 0$. Έστω ότι $\delta < 0$. Στην συνέχεια της απόδειξης συμβολίζουμε τους κοινούς δείκτες με k_n ενώ θα θεωρήσουμε ότι το $a_k \in [0,1]$ υπολογίζεται αναλυτικά (βέλτιστο βήμα).

- *(συνέχεια απόδειξης)*: Θεωρούμε την συνάρτηση $h(a) := f(x_{k_n} + a(y_{k_n} - x_{k_n}))$. Εύκολα υπολογίζουμε πως $h'(a) := \nabla f(x_{k_n} + a(y_{k_n} - x_{k_n}))^T(y_{k_n} - x_{k_n})$. Από το θεώρημα Μέσης Τιμής, υπάρχει ένα $\mu_a \in (0,a)$ ώστε να ισχύει
$h(a) - h(0) = a\nabla f(x_{k_n} + \mu_a(y_{k,n} - x_{k_n}))^T(y_{k_n} - x_{k_n})$
και επομένως καθώς
για $a \in (0,a'),\ k_n \geq k_{n_0},$ για κάποια a', k_{n_0} με $\epsilon_{k_n,a} \to 0$ όταν $k_n \to \infty,\ \ a \to 0$

$$\begin{aligned} f(x_{k_n} + a(y_{k_n} - x_{k_n})) - f(x_{k_n}) &= a\nabla f(x_{k_n} + \mu_a(y_{k,n} - x_{k_n}))^T(y_{k_n} - x_{k_n})) \\ &= a\left(\nabla f(\bar{x}) + \epsilon_{k_n,a}\right)^T\left((\bar{y} - \bar{x}) + \epsilon_{k_n,a}\right) \\ &= a(\delta + \epsilon_{k_n,a}) \\ &\leq a\delta\beta, \text{ για κάποιο } \beta \in (0,1) \end{aligned}$$

Από τον ορισμό του βέλτιστου βήματος έχουμε, (παρατηρούμε ότι $(\delta + \epsilon_{k_n,a}) < 0$ καθώς $\epsilon_{k_n,a} \to 0$),

$$\begin{aligned} f(x_{k_n} + a_k(y_{k_n} - (x_{k_n})) - f(x_{k_n}) &\leq f(x_{k_n} + a(y_{k_n} - (x_{k_n})) - f(x_{k_n}) \\ &\leq a\delta\beta, \text{ για κάποιο } \beta \in (0,1) \end{aligned}$$

Επειδή $\delta < 0$, και η ακολουθία $\{f(x_k)\}_{k=0,\ldots}$ είναι φθίνουσα, η παραπάνω σχέση οδηγεί στο συμπέρασμα πως $\lim_{k_n\to\infty} f(x_{k_n}) = -\infty$, που είναι άτοπο καθώς $\lim_{k_n\to\infty} f(x_{k_n}) = f(\bar{x})$ αλλά το $\bar{x} \in S$ (όπου S φραγμένο).

- Αν η $f \in \mathbb{R}^n \to \mathbb{R}$ είναι (επιπλέον των υποθέσεων του θεωρήματος) και <u>κυρτή</u>, τότε η μέθοδος υπολογίζει σημεία ολικού ελαχίστου.

- Αν η $f \in \mathbb{R}^n \to \mathbb{R}$ είναι (επιπλέον των υποθέσεων του θεωρήματος) και <u>αυστηρά κυρτή</u>, τότε η μέθοδος υπολογίζει το μοναδικό σημείο ελαχίστου και η ακολουθία $\{x_k\}_{k=1,..} \in S$ συγκλίνει σ'αυτό.

- Η μέθοδος εφαρμόζεται κυρίως όταν το σύνολο περιορισμών έχει απλή γεωμετρική μορφή, π.χ. είναι μία μπάλα, ένα (κυρτό) πολύεδρο κτλ. Τότε είναι εύκολο να υπολογιστεί αναλυτικά το βήμα 2 του αλγόριθμου, μέσω του θεωρήματος πολλαπλασιαστών Lagrange στην περίπτωση της μπάλας, η μέσω τεχνικών γραμμικού προγραμματισμού στην περίπτωση του πολύεδρου.

Παραδείγματα:

- Άσκηση: Έστω διανύσματα $x_k = [u, v, w]^T$, $\nabla f(x_k) = [a, b, c]^T$ και σύνολο $S = \{[x, y, z]^T \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq r^2\}$.
  Να υπολογιστεί το διάνυσμα του βήματος 2, του αλγόριθμου Frank-Wolfe δηλαδή, να λυθεί το πρόβλημα $\min_{y \in S} \nabla f(x_k)^T (y - x_k) = \nabla f(x_k)^T (y_k - x_k)$

- <u>Λύση</u>: Παρατηρούμε πως το υποσύνολο $S \subset \mathbb{R}^3$ είναι κλειστό και φραγμένο (άρα συμπαγές), ενώ η f είναι συνεχής. Επομένως από το Θεώρημα 1, υπάρχει τουλάχιστον ένα σημείο ελαχίστου. Επίσης το $S \subset \mathbb{R}^3$ είναι κυρτό, ενώ η συνάρτηση f είναι κυρτή. Επομένως αναζητούμε μόνο σημείο(α) ελαχίστου. Συμβολίζουμε με $\mathbf{y} = [x, y, z]^T$. Χρησιμοποιώντας το θεώρημα (KTL)

$$f_0(x, y, z) = \nabla f(x_k)^T (y - x_k) = ax - au + by - bv + cz - cw$$

$$f_1(x, y, z) = x^2 + y^2 + z^2 - r^2, l = m = 1,$$

έχουμε,

$$\nabla f_0(\bar{x}) + \lambda_1 \nabla f_1(\bar{x}) = 0,$$

$$\lambda_1 f_1(\bar{x}) = 0$$

$\lambda_1 \geq 0$ (αντιστοιχεί σε ανισότητα)

ή ισοδύναμα

- (*συνέχεια λύσης*):

$$\begin{bmatrix} a+2\lambda_1\bar{x} \\ b+2\lambda_1\bar{y} \\ c+2\lambda_1\bar{z} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\lambda_1(\bar{x}^2+\bar{y}^2+\bar{z}^2-r^2)=0, \quad \lambda_1 \geq 0$$

Περίπτωση 1: $\lambda_1 = 0$. Τότε έχουμε άτοπο καθώς $\nabla f(x_k) \neq 0^T$ και επομένως κάποια από τις συντεταγμένες δεν είναι μηδέν.

Περίπτωση 2: $\lambda_1 > 0$. Τότε,

$$\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} -\frac{a}{2\lambda_1} \\ -\frac{b}{2\lambda_1} \\ -\frac{c}{2\lambda_1} \end{bmatrix}$$

$$(\bar{x}^2+\bar{y}^2+\bar{z}^2-r^2)=0.$$

Αντικαθιστώντας τα x,y,z στην τελευταία σχέση, λαμβάνουμε

$$\frac{a^2+b^2+c^2}{4\lambda_1^2}=r^2, \quad \lambda_1=\frac{\sqrt{a^2+b^2+c^2}}{2r}, \bar{x}=-\frac{ra}{\|\nabla f(x_k)\|_2}, \bar{y}=-\frac{rb}{\|\nabla f(x_k)\|_2}, \bar{z}=-\frac{rc}{\|\nabla f(x_k)\|_2}.$$

- Παρατηρούμε πως καταλήξαμε σε ένα μόνο υποψήφιο σημείο ελαχίστου (Μοναδικό !) το οποίο βρίσκεται πάνω στο σύνορο, το σημείο, $y=-\frac{r}{\|\nabla f(x_k)\|_2}\nabla f(x_k)$

Ασκήσεις

- *Άσκηση:* Να χρησιμοποιήσετε την μέθοδο Frank-Wolfe, με αρχικό διάνυσμα $x_1 = [1/2, 1/2]^T$ για το πρόβλημα $\min_{x=[u,v]^T \in S} (u-3)^4 + (v-3)^4$ όπου $S = \{[u,v]^T \in \mathbb{R}^2 : u^2 + v^2 \leq 1\}$.

- *Λύση:* Παρατηρούμε πως η συνάρτηση $f(u,v) = (u-3)^4 + (v-3)^4$ είναι ομαλή και αυστηρά κυρτή, ενώ το σύνολο $S \subset \mathbb{R}^2$ είναι κυρτό και συμπαγές. Επομένως, υπάρχει μοναδικό σημείο ελαχίστου. Αναμένουμε πως ο αλγόριθμος θα συγκλίνει στο σημείο αυτό, καθώς ικανοποιούνται όλες οι υποθέσεις του θεωρήματος σύγκλισης. Υπολογίζουμε $\nabla f(u,v) = [4(u-3)^3, 4(v-3)^3]^T$. Αναζητούμε $y_1 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_1)^T (y - x_1) = \nabla f(x_1)^T (y_1 - x_1)$. Ορίζουμε συνάρτηση
$$h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_1)^T \left(\begin{bmatrix} u - 1/2 \\ v - 1/2 \end{bmatrix}\right) = -\frac{125}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \left(\begin{bmatrix} u - 1/2 \\ v - 1/2 \end{bmatrix}\right)$$
δηλαδή, $h(y) = h([u,v]^T) = -\dfrac{125}{2}(u+v) + \dfrac{125}{2}$.
Χρησιμοποιώντας το αποτέλεσμα της προηγούμενης άσκησης έχουμε
$$y_1 = -\frac{1}{\|\nabla f(x_1)\|_2} \nabla f(x_1) = -\frac{\sqrt{2}}{125} \begin{bmatrix} -\frac{125}{2} \\ -\frac{125}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

- <u>*Συνέχεια λύσης*</u>: Υπολογίζουμε $\delta_1 = \nabla f(x_1)^T(y_1 - x_1) = \begin{bmatrix} -\frac{125}{2} \\ -\frac{125}{2} \end{bmatrix}^T \begin{bmatrix} \frac{2-\sqrt{2}}{2\sqrt{2}} \\ \frac{2-\sqrt{2}}{2\sqrt{2}} \end{bmatrix} < 0.$

  Επομένως ο αλγόριθμος συνεχίζει. Υπολογισμός του $a_1 \in [0,1]$, το οποίο ορίζεται ως: $\min_{a \in [0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1))$. Ορίζουμε την συνάρτηση

  $$g(a) = f(x_1 + a(y_1 - x_1)) = f\left(\begin{bmatrix} (1-a)\frac{1}{2} + a\frac{1}{\sqrt{2}} \\ (1-a)\frac{1}{2} + a\frac{1}{\sqrt{2}} \end{bmatrix}\right) = 2\left((1-a)\frac{1}{2} + a\frac{1}{\sqrt{2}} - 3\right)^4$$

  Αναζητούμε, δηλαδή, το ελάχιστο της συνάρτησης στο [0,1]. Παρατηρούμε πως το ελάχιστο της συνάρτησης είναι το $a_1 = 1$. Επομένως έχουμε πως $x_2 = x_1 + a_1(y_1 - x_1) = y_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

  Παρατηρούμε πως ο υπολογισμός του $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T(y - x_2) = \nabla f(x_2)^T(y_2 - x_2)$ γίνεται με παρόμοιο τρόπο..

- *Συνέχεια λύσης*: Ορίζουμε συνάρτηση

$$h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_2)^T \left(\begin{bmatrix} u - \frac{1}{\sqrt{2}} \\ v - \frac{1}{\sqrt{2}} \end{bmatrix}\right) = -\frac{2}{\sqrt{2}}(3\sqrt{2}-1)^3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \left(\begin{bmatrix} u - \frac{1}{\sqrt{2}} \\ v - \frac{1}{\sqrt{2}} \end{bmatrix}\right)$$

  δηλαδή, $h(y) = h([u, v]^T) = -\dfrac{2}{\sqrt{2}}(3\sqrt{2}-1)^3(u+v) + c$

  Χρησιμοποιώντας τον τύπο της προηγούμενης άσκησης, λαμβάνουμε πως

$$y_2 = -\frac{1}{\|\nabla f(x_2)\|_2}\nabla f(x_2) = -\frac{1}{2(1-3\sqrt{2})^3}\begin{bmatrix} -\frac{2}{\sqrt{2}}(1-3\sqrt{2})^3 \\ -\frac{2}{\sqrt{2}}(1-3\sqrt{2})^3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

  Επομένως καταλήγουμε στην αναμενόμενη απάντηση $y_2 = x_2$ Εύκολα διαπιστώνουμε πως ο αλγόριθμος τερματίζει καθώς $\delta_2 = \nabla f(x_2)^T(y_2 - x_2) = 0$ και επομένως το x_2 είναι το μοναδικό σημείο ελαχίστου.

- Παρατηρήσεις: Υπολογίσαμε την παράμετρο επιτάχυνσης, αναλυτικά. Στην πράξη, χρησιμοποιούμε προσεγγιστικές διαδικασίες (π.χ τεχνικές γραμμικής αναζητήσεις). Η απλή γεωμετρία του χωρίου βοηθάει καθοριστικά στους υπολογισμούς.

- Το κύριο χαρακτηριστικό της μεθόδου Frank-Wolfe είναι η αναζήτηση ένος στοιχείου του συνόλου των περιορισμών (που ορίζει την κατεύθυνση της μεθόδου) το οποίο προκύπτει επιλύοντας ένα πρόβλημα ελαχιστοποίησης γραμμικής συνάρτησης πάνω στο σύνολο περιορισμών.

- *Άσκηση:* Να χρησιμοποιήσετε την μέθοδο Frank-Wolfe, με αρχικό διάνυσμα $x_1 = [1/4, 1/4]^T$ για το πρόβλημα $\min_{x=[u,v]^T \in S} (u-2)^2 + (v-2)^2$ όπου $S = \{[u,v]^T \in \mathbb{R}^2 : v \leq 1-u, u \geq 0, v \geq 0\}$.

- *Λύση:* Παρατηρούμε πως η συνάρτηση $f(u,v) = (u-2)^2 + (v-2)^2$ είναι ομαλή και αυστηρά κυρτή, ενώ το σύνολο $S \subset \mathbb{R}^2$ είναι κυρτό και συμπαγές (Άσκηση: Αποδείξτε ότι το σύνολο είναι κυρτό) . Επομένως, υπάρχει μοναδικό σημείο ελαχίστου. Αναμένουμε πως ο αλγόριθμος θα συγκλίνει στο σημείο αυτό, καθώς ικανοποιούνται όλες οι υποθέσεις του θεωρήματος σύγκλισης. Υπολογίζουμε $\nabla f(u,v) = [2(u-2), 2(v-2)]^T$. Αναζητούμε $y_1 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_1)^T (y - x_1) = \nabla f(x_1)^T (y_1 - x_1)$. Ορίζουμε συνάρτηση
$$h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_1)^T \left(\begin{bmatrix} u - 1/4 \\ v - 1/4 \end{bmatrix}\right) = -\frac{7}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \left(\begin{bmatrix} u - 1/4 \\ v - 1/4 \end{bmatrix}\right)$$
δηλαδή, $h(y) = h([u,v]^T) = -\frac{7}{2}(u+v) + \frac{7}{4}$.
Παρατηρούμε πως ελαχιστοποιούμε γραμμική συνάρτηση πάνω σε τρίγωνο. Επομένως τα υποψήφια σημεία ελαχίστου είναι οι κορυφές του τριγώνου δηλαδή, τα σημεία $[0,0]^T,\ [1,0]^T,\ [0,1]^T$. Παρατηρούμε πως τα σημεία $[1,0]^T,\ [0,1]^T$ είναι σημεία ελαχίστου.
Εξετάζουμε πρώτα το $y_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

- *Συνέχεια λύσης*: Υπολογίζουμε $\delta_1 = \nabla f(x_1)^T(y_1 - x_1) = \begin{bmatrix} -\frac{7}{2} \\ -\frac{7}{2} \end{bmatrix}^T \begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \end{bmatrix} < 0.$

Επομένως ο αλγόριθμος συνεχίζει. Υπολογισμός του $a_1 \in [0,1]$, το οποίο ορίζεται ως: $\min_{a \in [0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1))$. Ορίζουμε την συνάρτηση

$$g(a) = f(x_1 + a(y_1 - x_1)) = f\left(\begin{bmatrix} (1-a)\frac{1}{4} + a \\ (1-a)\frac{1}{4} \end{bmatrix}\right) = \left((1-a)\frac{1}{4} + a - 2\right)^2 + \left((1-a)\frac{1}{4} - 2\right)^2$$

Αναζητούμε, δηλαδή, το ελάχιστο της συνάρτησης στο [0,1]. Παρατηρούμε πως είναι το $a_1 = 1$. Προσεγγίζουμε το Επομένως έχουμε πως

$$x_2 = x_1 + a_1(y_1 - x_1) = y_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Παρατηρούμε πως ο υπολογισμός του $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T(y - x_2) = \nabla f(x_2)^T(y_2 - x_2)$ γίνεται με παρόμοιο τρόπο..

- Αναζητούμε $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T (y - x_2) = \nabla f(x_2)^T (y_2 - x_2)$. Ορίζουμε συνάρτηση $h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_2)^T \left(\begin{bmatrix} u-1 \\ v \end{bmatrix}\right) = \begin{bmatrix} -2 \\ -4 \end{bmatrix}^T \left(\begin{bmatrix} u-1 \\ v \end{bmatrix}\right)$ δηλαδή, $h(y) = h([u, v]^T) = -2u - 4v + 2$. Παρατηρούμε πως ελαχιστοποιούμε γραμμική συνάρτηση πάνω σε τρίγωνο. Επομένως τα υποψήφια σημεία ελαχίστου είναι οι κορυφές του τριγώνου δηλαδή, τα σημεία $[0,0]^T$, $[1,0]^T$, $[0,1]^T$. Παρατηρούμε πως το σημείο $[0,1]^T$ είναι σημεία ελαχίστου. Άρα $y_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

- Υπολογίζουμε $\delta_2 = \nabla f(x_2)^T(y_2 - x_2) = \begin{bmatrix} -2 \\ -4 \end{bmatrix}^T \begin{bmatrix} -1 \\ 1 \end{bmatrix} < 0.$

  Επομένως ο αλγόριθμος συνεχίζει. Υπολογισμός του $a_2 \in [0,1]$, το οποίο ορίζεται ως: $\min_{a \in [0,1]} f(x_2 + a(y_2 - x_2)) = f(x_2 + a_2(y_2 - x_2))$. Ορίζουμε την συνάρτηση

  $$g(a) = f(x_2 + a(y_2 - x_2)) = f\left(\begin{bmatrix} (1-a) \\ \alpha \end{bmatrix}\right) = \big((1-a) - 2\big)^2 + (\alpha - 2)^2$$

  Αναζητούμε, δηλαδή, το ελάχιστο της συνάρτησης στο [0,1]. Παρατηρούμε πως είναι το $a_2 = 1/2$. Προσεγγίζουμε το Επομένως έχουμε πως

  $$x_3 = x_2 + a_2(y_2 - x_2) = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}.$$

- Παρατηρήσεις: Εύκολα διαπιστώνουμε πως το $\delta_3 = 0$ και ο αλγόριθμος σταματά έχοντας υπολογίσει το μοναδικό σημείο ελαχίστου.

- Αν χρησιμοποιούσαμε στο πρώτο βήμα ως διάνυσμα $y_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ θα καταλήγαμε (μετά από παρόμοιες πράξεις στο ίδιο αποτέλεσμα.

- *Άσκηση:* Να χρησιμοποιήσετε την μέθοδο Frank-Wolfe, με αρχικό διάνυσμα $x_1 = [1/4, 1/4]^T$ για το πρόβλημα $\min_{x=[u,v]^T \in S} u^3 - v^3$ όπου $S = \{[u,v]^T \in \mathbb{R}^2 : v \leq 1 - u, u \geq 0, v \geq 0\}$.

- *Λύση:* Παρατηρούμε πως η συνάρτηση $f(u,v) = u^3 - v^3$ είναι ομαλή, ενώ το σύνολο $S \subset \mathbb{R}^2$ είναι κυρτό και συμπαγές. Επομένως, υπάρχει σημείο ελαχίστου. Παρατηρούμε πως η συνάρτηση δεν είναι κυρτή. Οι υποθέσεις του θεωρήματος σύγκλισης ικανοποιούνται. Υπολογίζουμε $\nabla f(u,v) = [3u^2, -3v^2]^T$. Αναζητούμε $y_1 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_1)^T (y - x_1) = \nabla f(x_1)^T (y_1 - x_1)$. Ορίζουμε συνάρτηση

$$h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_1)^T \left(\begin{bmatrix} u - 1/4 \\ v - 1/4 \end{bmatrix}\right) = \frac{3}{4^2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} u - 1/4 \\ v - 1/4 \end{bmatrix}\right)$$

δηλαδή, $h(y) = h([u,v]^T) = \dfrac{3}{16}(u - v)$.

Παρατηρούμε πως ελαχιστοποιούμε γραμμική συνάρτηση πάνω σε τρίγωνο. Επομένως τα υποψήφια σημεία ελαχίστου είναι οι κορυφές του τριγώνου δηλαδή, τα σημεία $[0,0]^T$, $[1,0]^T$, $[0,1]^T$. Παρατηρούμε πως το σημείο ελαχίστου είναι το $[0,1]^T$.

- <u>*Συνέχεια λύσης*</u>: Υπολογίζουμε $\delta_1 = \nabla f(x_1)^T(y_1 - x_1) = \begin{bmatrix} \frac{3}{16} \\ -\frac{3}{16} \end{bmatrix}^T \begin{bmatrix} -\frac{1}{4} \\ \frac{3}{4} \end{bmatrix} < 0.$

Επομένως ο αλγόριθμος συνεχίζει. Υπολογισμός του $a_1 \in [0,1]$, το οποίο ορίζεται ως: $\min_{a \in [0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1))$. Ορίζουμε την συνάρτηση

$$g(a) = f(x_1 + a(y_1 - x_1)) = f\left(\begin{bmatrix} (1-a)\frac{1}{4} \\ (1-a)\frac{1}{4} + a \end{bmatrix}\right) = \left((1-a)\frac{1}{4}\right)^3 - \left((1-a)\frac{1}{4} + a\right)^3$$

$$= -a\left((1-a)^2\frac{1}{16} + (1-a)\frac{1}{4}((1-a)\frac{1}{4} + a) + ((1-a)\frac{1}{4} + a)^2\right)$$

Αναζητούμε, δηλαδή, το ελάχιστο της συνάρτησης στο [0,1]. Παρατηρούμε πως είναι το $a_1 = 1$. Προσεγγίζουμε το Επομένως έχουμε πως $x_2 = x_1 + a_1(y_1 - x_1) = y_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Παρατηρούμε πως ο υπολογισμός του $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T(y - x_2) = \nabla f(x_2)^T(y_2 - x_2)$ γίνεται με παρόμοιο τρόπο..

- Αναζητούμε $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T(y - x_2) = \nabla f(x_2)^T(y_2 - x_2)$. Ορίζουμε συνάρτηση $h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_2)^T\left(\begin{bmatrix} u \\ v-1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}^T\left(\begin{bmatrix} u \\ v-1 \end{bmatrix}\right)$ δηλαδή, $h(y) = h([u, v]^T) = -v + 1$.
  Παρατηρούμε πως ελαχιστοποιούμε γραμμική συνάρτηση πάνω σε τρίγωνο. Επομένως τα υποψήφια σημεία ελαχίστου είναι οι κορυφές του τριγώνου δηλαδή, τα σημεία $[0,0]^T,\ [1,0]^T,\ [0,1]^T$. Παρατηρούμε πως το σημείο $[0,1]^T$ είναι σημεία ελαχίστου. Άρα $y_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Υπολογίζουμε $\delta_2 = \nabla f(x_2)^T(y_2 - x_2) = 0$ άρα ο αλγόριθμος τερματίζει. Έχουμε βρει το σημείο $\bar{x} := x_2 = [0,1]^T$ ως υποψήφιο σημείο ελαχίστου. Παρατηρούμε πως είναι το αναμενόμενο σημείο ελαχίστου!

- *Άσκηση:* Να χρησιμοποιήσετε την μέθοδο Frank-Wolfe, με αρχικό διάνυσμα $x_1 = [1/4, 1/4]^T$ για το πρόβλημα $\min_{x=[u,v]^T \in S} u^3 - v^3$ όπου $S = \{[u, v]^T \in \mathbb{R}^2 : u^2 + v^2 \leq 1\}$.

- *Λύση:* Παρατηρούμε πως η συνάρτηση $f(u, v) = u^3 - v^3$ είναι ομαλή, ενώ το σύνολο $S \subset \mathbb{R}^2$ είναι κυρτό και συμπαγές. Επομένως, υπάρχει σημείο ελαχίστου. Παρατηρούμε πως η συνάρτηση δεν είναι κυρτή. Οι υποθέσεις του θεωρήματος σύγκλισης ικανοποιούνται. Υπολογίζουμε $\nabla f(u, v) = [3u^2, -3v^2]^T$. Αναζητούμε $y_1 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_1)^T (y - x_1) = \nabla f(x_1)^T (y_1 - x_1)$. Ορίζουμε συνάρτηση

$$h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_1)^T \left(\begin{bmatrix} u - 1/4 \\ v - 1/4 \end{bmatrix}\right) = \frac{3}{4^2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} u - 1/4 \\ v - 1/4 \end{bmatrix}\right)$$

δηλαδή, $h(y) = h([u, v]^T) = \dfrac{3}{16}(u - v)$.

Παρατηρούμε πως ελαχιστοποιούμε γραμμική συνάρτηση πάνω στην μοναδιαία σφαίρα. Επομένως εύκολα υπολογίζουμε,

$$y_1 = -\frac{1}{\|\nabla f(x_1)\|_2} \nabla f(x_1) = -\frac{1}{\sqrt{2}\frac{9}{16}} \begin{bmatrix} \frac{9}{16} \\ -\frac{9}{16} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

- <u>*Συνέχεια λύσης*</u>: Υπολογίζουμε $\delta_1 = \nabla f(x_1)^T(y_1 - x_1) = \begin{bmatrix} \frac{3}{16} \\ -\frac{3}{16} \end{bmatrix}^T \begin{bmatrix} -\frac{1}{\sqrt{2}} - \frac{1}{4} \\ \frac{1}{\sqrt{2}} - \frac{1}{4} \end{bmatrix} < 0.$

  Επομένως ο αλγόριθμος συνεχίζει. Υπολογισμός του $a_1 \in [0,1]$, το οποίο ορίζεται ως: $\min_{a \in [0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1))$. Ορίζουμε την συνάρτηση

  $$g(a) = f(x_1 + a(y_1 - x_1)) = f\left(\begin{bmatrix} (1-a)\frac{1}{4} - \frac{1}{\sqrt{2}}a \\ (1-a)\frac{1}{4} + \frac{1}{\sqrt{2}}a \end{bmatrix}\right) = \left((1-a)\frac{1}{4} - \frac{1}{\sqrt{2}}a\right)^3 - \left((1-a)\frac{1}{4} + \frac{1}{\sqrt{2}}a\right)^3$$

  $$= -\frac{2}{\sqrt{2}}a\left(3(1-a)^2\frac{1}{16} + \frac{1}{2}\alpha^2\right)$$

  Αναζητούμε, δηλαδή, το ελάχιστο της συνάρτησης στο [0,1]. Παρατηρούμε πως είναι το $a_1 = 1$.

  Προσεγγίζουμε το Επομένως έχουμε πως $x_2 = x_1 + a_1(y_1 - x_1) = y_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

  Παρατηρούμε πως ο υπολογισμός του $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T(y - x_2) = \nabla f(x_2)^T(y_2 - x_2)$ γίνεται με παρόμοιο τρόπο.

- Αναζητούμε $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T (y - x_2) = \nabla f(x_2)^T (y_2 - x_2)$. Ορίζουμε συνάρτηση

$$h : S \to \mathbb{R},\ h(y) = h\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \nabla f(x_2)^T \left(\begin{bmatrix} u + \frac{1}{\sqrt{2}} \\ v - \frac{1}{\sqrt{2}} \end{bmatrix}\right) = \begin{bmatrix} \frac{3}{2} \\ -\frac{3}{2} \end{bmatrix}^T \left(\begin{bmatrix} u + \frac{1}{\sqrt{2}} \\ v - \frac{1}{\sqrt{2}} \end{bmatrix}\right)$$

δηλαδή, $h(y) = h([u,v]^T) = \dfrac{3}{2}(u - v) + \dfrac{3}{\sqrt{2}}$.

Παρατηρούμε πως ελαχιστοποιούμε γραμμική συνάρτηση πάνω στην μοναδιαία σφαίρα. Επομένως εύκολα υπολογίζουμε,

$$y_2 = -\frac{1}{\|\nabla f(x_2)\|_2} \nabla f(x_2) = -\frac{1}{\sqrt{2}\frac{3}{2}} \begin{bmatrix} \frac{3}{2} \\ -\frac{3}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

. Υπολογίζουμε

$\delta_2 = \nabla f(x_2)^T (y_2 - x_2) = 0$ άρα ο αλγόριθμος τερματίζει. Έχουμε βρει το σημείο $\bar{x} := x_2$ που σύμφωνα με το θεώρημα ικανοποιεί την αναγκαία συνθήκη. ***Είναι σημείο ελαχίστου ?***

Όχι…παρατηρούμε πως $-1 = f(0,1) \leq f(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = -\frac{1}{\sqrt{2}}$.

Η μέθοδος των προβεβλημένων κλίσεων - Projected Gradient

- Υπενθύμιση Πρόβλημα (ΥΠΠ): Να υπολογιστούν σημεία τοπικού ή / και ολικού ελάχιστου για το πρόβλημα $\min_{x \in S} f(x) = f(\bar{x})$.

- Υποθέτουμε ότι η συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$, ικανοποιεί $f \in C^1(\mathbb{R}^n)$, ενώ το σύνολο $S \neq \emptyset, S \subset \mathbb{R}^n$ κυρτό και συμπαγές. Επομένως από το Θεώρημα 1 υπάρχει τουλάχιστον ένα σημείο ελαχίστου. Η ιδιότητα του κυρτού υποσυνόλου $S \subset \mathbb{R}^n$ είναι εξαιρετικά σημαντικός στην κατασκευή των αριθμητικών μεθόδων.

- Αναγκαία συνθήκη: $\nabla f(\bar{x})^T (x - \bar{x}) \geq 0, \quad \forall x \in S$.

- Οι υπολογιστικές μέθοδοι, για δοσμένη αρχική συνθήκη $x_0 \in \mathbb{R}^n$, κατασκευάζουν ακολουθία $\{x_k\}_{k=0}^{\infty} \in \mathbb{R}^n$ που (στο όριο - με κατάλληλο τρόπο) προσεγγίζει τη λύση ή λύσεις του προβλήματος (ΥΠΠ), δηλαδή λύσεις που ικανοποιούν την αναγκαία συνθήκη.

- Ο αλγόριθμος των προβεβλημένων κλίσεων (Projected Gradient Method).
- Με Χρήση των “μετρητών” $\delta_k := \nabla f(x_k)^T(y_k - x_k)$,

  $\zeta_k := \nabla f(x_k)^T(y_k - x_k) + \frac{\gamma}{2}\|y_k - x_k\|_2^2$

  1. Set $k = 0$. Given $x_0 \in S$, $\gamma > 0$.

     For $k = 1,..,$ compute until convergence

  2. Compute $y_k \in S$ such that

  $$\min_{y \in S}\left(\nabla f(x_k)^T(y - x_k) + \frac{\gamma}{2}\|y - x_k\|_2^2 \right)$$

  $$= \nabla f(x_k)^T(y_k - x_k) + \frac{\gamma}{2}\|y_k - x_k\|_2^2 := \zeta_k$$

  3. Compute $\delta_k = \nabla f(x_k)^T(y_k - x_k)$ If $\delta_k = 0$ Stop. Else
  4. Compute a_k such that

  $$\min_{a \in [0,1]} f(x_k + a(y_k - x_k)) = f(x_k + a_k(y_k - x_k))$$

  5. $x_{k+1} = x_k + a_k(y_k - x_k)$

- Παρατήρηση: Το διάνυσμα $y_k \in S$ που υπολογίζεται στο βήμα 2, είναι η προβολή του διανύσματος $z_k := x_k - \frac{1}{\gamma}\nabla f(x_k)$ πάνω στο σύνολο S, και συμβολίζεται ως $y_k = P_S z_k$.

- Απόδειξη:

$$\|y - z_k\|_2^2 = \left(y - (x_k - \frac{1}{\gamma}\nabla f(x_x))\right)^T \left(y - (x_k - \frac{1}{\gamma}\nabla f(x_x))\right)$$

$$= (y - x_k)^T(y - x_k) + \frac{2}{\gamma}(y - x_k)^T \nabla f(x_k) + \frac{1}{\gamma^2}\nabla f(x_k)^T \nabla f(x_k)$$

$$= \|y - x_k\|_2^2 + \frac{2}{\gamma}(y - x_k)^T \nabla f(x_k) + \frac{1}{\gamma^2}\|\nabla f(x_k)\|_2^2$$

$$= \frac{2}{\gamma}\left(\frac{\gamma}{2}\|y - x_k\|_2^2 + \nabla f(x_k)^T(y - x_k) + \frac{\gamma}{2}\|\nabla f(x_k)\|_2^2\right)$$

Παρατηρούμε πως ο τελευταίος όρος είναι ανεξάρτητος του $y \in S$. Επομένως, η ελαχιστοποίηση της ποσότητας $\nabla f(x_k)^T(y - x_k) + \frac{\gamma}{2}\|y - x_k\|^2$ πάνω στο S είναι ισοδύναμη με την ελαχιστοποίηση της απόστασης (ως προς την 2-νόρμα) από το $z_k := x_k - \frac{1}{\gamma}\nabla f(x_k)$.

- *Θεώρημα:* Έστω $f \in \mathbb{R}^n \to \mathbb{R}, f \in C^1(\mathbb{R}^n)$. Υποθέτουμε ότι το σύνολο $S \subset \mathbb{R}^n$ είναι κυρτό και συμπαγές. Αν ο αλγόριθμος σταματά για κάποιο κ, τότε $\nabla f(x_k)^T(y - x_k) \geq 0, \ \forall y \in S$ δηλαδή το $\bar{x} := x_k$ ικανοποιεί την αναγκαία συνθήκη. Διαφορετικά, η ακολουθία $\{x_k\}_{k=0,\ldots}$ που παράγει ο αλγόριθμος είναι άπειρη και κάθε όριο $\bar{x}$ συγκλίνουσας υπακολουθίας ικανοποιεί $\nabla f(\bar{x})^T(y - \bar{x}) \geq 0, \ \forall \ y \in S$ και ολόκληρες οι ακολουθίες $\delta_k \to 0, \ \zeta_k \to 0$.

- *Απόδειξη:* Θα δείξουμε επαγωγικά ότι αν $\delta_k \neq 0$, η ακολουθία $\{x_k\}_{k=0,\ldots}$,είναι καλά ορισμένη, (δηλαδή υπάρχει το $x_k \in S,$ για κάθε κ) και ότι η $\{f(x_k)\}_{k=0,\ldots}$ είναι φθίνουσα. Έστω $x_1 \in S$. Ορίζουμε $y_0 = x_1$. Παρατηρούμε πως το πρόβλημα του βήματος 2, είναι καλά ορισμένο καθώς η συνάρτηση $y \in S \to \nabla f(x_1)^T(y - x_1) + \dfrac{\gamma}{2}\|y - x_1\|_2^2 \in \mathbb{R}$ είναι συνεχής (τετραγωνική ως προς y) ενώ το $S \subset \mathbb{R}^n$ είναι συμπαγές. Επομένως υπάρχει τουλάχιστον ένα σημείο ελαχίστου y_1 στο S , δηλαδή, $\min\limits_{y \in S}\left(\nabla f(x_1)^T(y - x_1) + \dfrac{\gamma}{2}\|y - x_1\|_2^2 \right) = \nabla f(x_1)^T(y_1 - x_1) + \dfrac{\gamma}{2}\|y_1 - x_1\|_2^2$. Παρατηρούμε πως η συνάρτηση $y \in S \to \nabla f(x_1)^T(y - x_1) + \dfrac{\gamma}{2}\|y - x_1\|_2^2 \in \mathbb{R}$ είναι αυστηρά κυρτή, ορισμένη πάνω σε κυρτό σύνολο, επομένως το ελάχιστο y_1 είναι *μοναδικό σημείο (ολικού) ελαχίστου*. Εύκολα διαπιστώνουμε πως το βήμα 4 είναι επίσης καλά ορισμένο καθώς πρόκειται για πρόβλημα ελαχιστοποίησης συνεχούς συνάρτησης ορισμένης στο συμπαγές [0,1]. Επίσης επειδή τα $x_1, y_1 \in S$ και $a_1 \in [0,1]$, έχουμε ότι το $x_2 = x_1 + a_1(y_1 - x_1) = (1 - a_1)x_1 + a_1 y_1 \in S$. ($S \subset \mathbb{R}^n$ κυρτό).

- (Συνέχεια απόδειξης:) Παρατηρούμε επίσης, από τον ορισμό του a_1 πως,

$$\min_{a\in[0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1)) = f(x_2) \leq f(x_1 + 0(y_1 - x_1))$$

δηλαδή, $f(x_2) \leq f(x_1)$.
Επαγωγικά, χρησιμοποιώντας παρόμοια επιχειρήματα, αποδεικνύεται πως η ακολουθία $\{x_k\}_{k=1,\ldots}$,είναι καλά ορισμένη, (δηλαδή υπάρχει το $x_k \in S$, για κάθε κ) και ότι η $\{f(x_k)\}_{k=1,\ldots}$ είναι φθίνουσα. Προφανώς αν ο αλγόριθμος σταματά για κάποιο $\delta_k := \nabla f(x_k)^T(y_k - x_k) = 0$ το $\bar{x} := x_k$ ικανοποιεί την αναγκαία συνθήκη καθώς,

$$0 = \nabla f(x_k)^T(y_k - x_k) \leq \nabla f(x_k)^T(y_k - x_k) + \frac{\gamma}{2}\|y_k - x_k\|_2^2$$

$$:= \min_{y\in S} \nabla f(x_k)^T(y - x_k) + \frac{\gamma}{2}\|y - x_k\|_2^2$$

Διαλέγουμε το $x_k + \theta(y - x_k) = (1-\theta)x_k + \theta y,\ \theta \in (0,1]\ \ x_k, y \in S$. Παρατηρούμε πως το $x_k + \theta(y - x_k) \in S$ επειδή το S είναι κυρτό.
Επομένως, έχουμε,

$$0 \leq \nabla f(x_k)^T(x_k + \theta(y - x_k) - x_k) + \frac{\gamma}{2}\|x_k + \theta(y - x_k) - x_k\|_2^2$$

$$= \theta \nabla f(x_k)^T(y - x_k) + \theta^2\|y - x_k\|_2^2$$

Διαρώντας με $\theta \in (0,1]$, $\theta \to 0$, λαμβάνουμε ότι $\nabla f(x_k)^T(y - x_k) \geq 0,\ \forall y \in S$.

- *(Συνέχεια Απόδειξης):* Επομένως η ακολουθία είναι άπειρη. Επειδή το $S \subset \mathbb{R}^n$είναι συμπαγές και ολόκληρη η ακολουθία $x_k, y_k \in S,$ για κάθε κ, έχουμε πως υπάρχει συγκλίνουσα υπακολουθία, δηλαδή υπάρχει $\bar{x} \in S$, $\bar{y} \in S$ ώστε $\lim_{k_n \to \infty} x_{k_n} = \bar{x}$ και $\lim_{k_l \to \infty} y_{k_l} = \bar{y}$.

  Διαλέγοντας του κοινούς δείκτες και επειδή, $f \in C^1(\mathbb{R}^n)$, και $\zeta_{k_{n,l}} \leq 0,$

$$\zeta_{k_{n,l}} = \min_{y \in S} \left(\nabla f(x_{k_{n,l}})^T (y - x_{k_{n,l}}) + \frac{\gamma}{2} \|y - x_{n_{k,l}}\|_2^2 \right)$$

$$\leq \nabla f(x_{k_{n,l}})^T (x_{k_{n,l}} - x_{k_{n,l}}) + \frac{\gamma}{2} \|x_{k_{n,l}} - x_{k_{n,l}}\|_2^2 = 0$$

$$\zeta_{k_{n,l}} = \nabla f(x_{k_{n,l}})^T (y_{k_{n,l}} - x_{k_{n,l}}) + \frac{\gamma}{2} \|y_{k_{n,l}} - x_{k_{n,l}}\|_2^2$$

$$\to \nabla f(\bar{x})^T (\bar{y} - \bar{x}) + \frac{\gamma}{2} \|\bar{y} - \bar{x}\|_2^2 := \zeta \leq 0$$

  Επομένως, έχουμε ότι $\nabla f(\bar{x})^T (\bar{y} - \bar{x}) + \dfrac{\gamma}{2} \|\bar{y} - \bar{x}\|_2^2 \leq 0$ ή ισοδύναμα $\delta := \nabla f(\bar{x})^T (\bar{y} - \bar{x}) \leq 0$. Θα δείξουμε ότι $\delta = 0.$ Έστω ότι $\delta < 0.$ Στην συνέχεια της απόδειξης συμβολίζουμε τους κοινούς δείκτες με k_n ενώ θα θεωρήσουμε ότι το $a_k \in [0,1]$ υπολογίζεται αναλυτικά (βέλτιστο βήμα).

- (*συνέχεια απόδειξης*): Θεωρούμε την συνάρτηση $h(a) := f(x_{k_n} + a(y_{k_n} - x_{k_n}))$. Εύκολα υπολογίζουμε πως $h'(a) := \nabla f(x_{k_n} + a(y_{k_n} - x_{k_n}))^T(y_{k_n} - x_{k_n})$. Από το θεώρημα Μέσης Τιμής, υπάρχει ένα $\mu_a \in (0,a)$ ώστε να ισχύει
$h(a) - h(0) = a\nabla f(x_{k_n} + \mu_a(y_{k,n} - x_{k_n}))^T(y_{k_n} - x_{k_n})$
και επομένως καθώς
για $a \in (0,a'),\ k_n \geq k_{n_0}$, για κάποια a', k_{n_0} με $\epsilon_{k_n,a} \to 0$ όταν $k_n \to \infty,\ \ a \to 0$

$$
\begin{aligned}
f(x_{k_n} + a(y_{k_n} - x_{k_n})) - f(x_{k_n}) &= a\nabla f(x_{k_n} + \mu_a(y_{k,n} - x_{k_n}))^T(y_{k_n} - x_{k_n})) \\
&= a\left(\nabla f(\bar{x}) + \epsilon_{k_n,a}\right)^T\left((\bar{y} - \bar{x}) + \epsilon_{k_n,a}\right) \\
&= a(\delta + \epsilon_{k_n,a}) \\
&\leq a\delta\beta, \text{ για κάποιο } \beta \in (0,1)
\end{aligned}
$$

Από τον ορισμό του βέλτιστου βήματος έχουμε, (παρατηρούμε ότι $(\delta + \epsilon_{k_n,a}) < 0$ καθώς $\epsilon_{k_n,a} \to 0$),

$$
\begin{aligned}
f(x_{k_n} + a_k(y_{k_n} - (x_{k_n})) - f(x_{k_n}) &\leq f(x_{k_n} + a(y_{k_n} - (x_{k_n})) - f(x_{k_n}) \\
&\leq a\delta\beta, \text{ για κάποιο } \beta \in (0,1)
\end{aligned}
$$

Επειδή $\delta < 0$, και η ακολουθία $\{f(x_k)\}_{k=0,\ldots}$ είναι φθίνουσα, η παραπάνω σχέση οδηγεί στο συμπέρασμα πως $\lim_{k_n \to \infty} f(x_{k_n}) = -\infty$, που είναι άτοπο καθώς $\lim_{k_n \to \infty} f(x_{k_n}) = f(\bar{x})$ αλλά το $\bar{x} \in S$ (όπου S φραγμένο).

- *Συνέχεια απόδειξης:* Επομένως έχουμε αποδείξει ότι $\delta = 0$. Παρατηρούμε όμως ότι $\delta \leq \zeta \leq 0$, άρα έχουμε επίσης ότι $\zeta = 0$.
  Από τον ορισμό του y_k (βήμα 2), έχουμε
  $$\nabla f(x_{k_{n,l}})^T(y - x_{k_{n,l}}) + \frac{\gamma}{2}\|y - x_{k_{n,l}}\|_2^2 \geq \nabla f(x_{k_{n,l}})^T(y_{k_{n,l}} - x_{k_{n,l}}) + \frac{\gamma}{2}\|y_{k_{n,l}} - x_{k_{n,l}}\|_2^2 := \zeta_{k_{n,l}}$$
  Παίρνοντας το όριο καταλήγουμε, για κάθε $y \in S$,
  $$\nabla f(\bar{x})^T(y - \bar{x}) + \frac{\gamma}{2}\|y - \bar{x}\|_2^2 \geq \zeta := 0.$$
  Εφαρμόζουμε το θ- επιχείρημα. Διαλέγουμε το
  $\bar{x} + \theta(y - \bar{x}) = (1 - \theta)\bar{x} + \theta y,\ \theta \in (0,1]\ \ \bar{x}, y \in S$. Παρατηρούμε πως το $\bar{x} + \theta(y - \bar{x}) \in S$ επειδή το S είναι κυρτό.
  Επομένως, έχουμε,
  $$0 \leq \nabla f(\bar{x})^T(\bar{x} + \theta(y - \bar{x}) - \bar{x}) + \frac{\gamma}{2}\|\bar{x} + \theta(y - \bar{x}) - \bar{x}\|_2^2$$
  $$= \theta \nabla f(\bar{x})^T(y - \bar{x}) + \theta^2\|y - \bar{x}\|_2^2$$
  Διαιρώντας με $\theta \in (0,1],\ \theta \to 0$, λαμβάνουμε ότι $\nabla f(\bar{x})^T(y - \bar{x}) \geq 0,\ \forall y \in S$.

- *Άσκηση*: Δείξτε ότι ολόκληρες οι ακολουθίες $\delta_k, \zeta_k \to 0$.

- *Άσκηση:* Έστω διανύσματα x_k, $\nabla f(x_k)$ τα διανύσματα που παράγονται στην κ-επανάληψη του αλγόριθμου των προβεβλημένων κλίσεων με σύνολο περιορισμών $S = \{[x,y,z]^T \in \mathbb{R}^3 : x^2+y^2+z^2 \leq r^2\}$.
Να υπολογιστεί το διάνυσμα του βήματος 2, του αλγόριθμου προβεβλημένων κλίσεων δηλαδή, να λυθεί το πρόβλημα $\min_{y\in S} \|y - z_k\|_2^2$ όπου $z_k = x_k - \frac{1}{\gamma}\nabla f(x_k)$.

- Λύση: Παρατηρούμε πως το υποσύνολο $S \subset \mathbb{R}^3$ είναι κλειστό και φραγμένο (άρα συμπαγές), ενώ η f είναι συνεχής. Επομένως από το Θεώρημα 1, υπάρχει τουλάχιστον ένα σημείο ελαχίστου. Επίσης το $S \subset \mathbb{R}^3$ είναι κυρτό, ενώ η συνάρτηση f είναι αυστηρά κυρτή. Επομένως αναζητούμε μοναδικό σημείο ελαχίστου. Συμβολίζουμε με $\mathbf{y} = [x,y,z]^T$, και $z = [u,v,w]^T$.
Δεν χρειάζεται να υπολογίσουμε το διάνυσμα! Χρησιμοποιώντας το θεώρημα (KTL)
$$f_0(x,y,z) = \|\mathbf{y} - u\|_2^2 = (x-v)^2 + (y-v)^2 + (z-w)^2$$
$$f_1(x,y,z) = x^2 + y^2 + z^2 - r^2, l = m = 1,$$
έχουμε,
$$\nabla f_0(\bar{x}) + \lambda_1 \nabla f_1(\bar{x}) = 0,$$
$$\lambda_1 f_1(\bar{x}) = 0$$
$\lambda_1 \geq 0$ (αντιστοιχεί σε ανισότητα)
ή ισοδύναμα

- (*συνέχεια λύσης*):

$$\begin{bmatrix} 2(\bar{x}-u)+2\lambda_1\bar{x} \\ 2(\bar{y}-v)+2\lambda_1\bar{y} \\ 2(\bar{z}-w)+2\lambda_1\bar{z} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\lambda_1(\bar{x}^2+\bar{y}^2+\bar{z}^2-r^2)=0, \quad \lambda_1 \geq 0$$

Περίπτωση 1: $\lambda_1 = 0$. Τότε έχουμε άτοπο καθώς $\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}$ και επομένως έχουμε ότι αν

$z \in S,$ δηλαδή $u^2+v^2+w^2 \leq r^2$, τότε έχουμε, $P_s z = z$.

Περίπτωση 2: $\lambda_1 > 0$. Τότε,

$$\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} \frac{u}{(1+\lambda_1)} \\ \frac{v}{(1+\lambda_1)} \\ \frac{w}{(1+\lambda_1)} \end{bmatrix}$$

$$(\bar{x}^2+\bar{y}^2+\bar{z}^2-r^2)=0.$$

Αντικαθιστώντας τα x,y,z στην τελευταία σχέση, λαμβάνουμε

$$\frac{u^2+v^2+w^2}{(1+\lambda_1)^2}=r^2, \quad (1+\lambda_1)=\frac{\sqrt{u^2+v^2+w^2}}{r}, \bar{x}=\frac{ru}{\|z\|_2}, \bar{y}=\frac{rv}{\|z\|_2}, \bar{z}=\frac{rw}{\|z\|_2}.$$

- Παρατηρούμε πως καταλήξαμε σε ένα μόνο υποψήφιο σημείο ελαχίστου (Μοναδικό !) το οποίο βρίσκεται πάνω στο σύνορο, το σημείο, $y = \frac{r}{\|z\|_2} z$

- Να χρησιμοποιήσετε την μέθοδο των προβεβλημένων κλίσεων, με αρχικό διάνυσμα $x_1 = [1/2, 1/2]^T$ και $\gamma = 1$, για το πρόβλημα $\min_{x=[u,v]^T \in S} (u-3)^4 + (v-3)^4$ όπου $S = \{[u,v]^T \in \mathbb{R}^2 : u^2 + v^2 \leq 1\}$.

- *Λύση:* Παρατηρούμε πως η συνάρτηση $f(u,v) = (u-3)^4 + (v-3)^4$ είναι ομαλή και αυστηρά κυρτή, ενώ το σύνολο $S \subset \mathbb{R}^2$ είναι κυρτό και συμπαγές. Επομένως, υπάρχει μοναδικό σημείο ελαχίστου. Αναμένουμε πως ο αλγόριθμος θα συγκλίνει στο σημείο αυτό, καθώς ικανοποιούνται όλες οι υποθέσεις του θεωρήματος σύγκλισης. Υπολογίζουμε $\nabla f(u,v) = [4(u-3)^3, 4(v-3)^3]^T$. Αναζητούμε $y_1 \in S$ τέτοιο ώστε $\min_{y \in S}\left(\nabla f(x_1)^T(y - x_1) + \frac{\gamma}{2}\|y - x_1\|_2^2\right) = \nabla f(x_1)^T(y_1 - x_1) + \frac{\gamma}{2}\|y_1 - x_1\|_2^2$. Ορίζουμε διάνυσμα $z_1 = x_1 - \frac{1}{\gamma}\nabla f(x_1) := \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} - \frac{1}{1}\begin{bmatrix} -\frac{125}{2} \\ -\frac{125}{2} \end{bmatrix} = \begin{bmatrix} \frac{126}{2} \\ \frac{126}{2} \end{bmatrix}$. Επομένως αναζητούμε την προβολή του διανύσματος, $y_1 = P_s z_1$. Εύκολα αποδεικνύεται, από το αποτέλεσμα της προηγούμενης άσκησης έχουμε

$$y_1 = \frac{1}{\|z_1\|_2} z_1 = \frac{\sqrt{2}}{126}\begin{bmatrix} \frac{126}{2} \\ \frac{126}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

- <u>*Συνέχεια λύσης*</u>: Υπολογίζουμε $\delta_1 = \nabla f(x_1)^T(y_1 - x_1) = \begin{bmatrix} -\frac{125}{2} \\ -\frac{125}{2} \end{bmatrix}^T \begin{bmatrix} \frac{2-\sqrt{2}}{2\sqrt{2}} \\ \frac{2-\sqrt{2}}{2\sqrt{2}} \end{bmatrix} < 0.$

  Επομένως ο αλγόριθμος συνεχίζει. Υπολογισμός του $a_1 \in [0,1]$, το οποίο ορίζεται ως: $\min_{a\in[0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1))$. Ορίζουμε την συνάρτηση

$$g(a) = f(x_1 + a(y_1 - x_1)) = f\left(\begin{bmatrix} (1-a)\frac{1}{2} + a\frac{1}{\sqrt{2}} \\ (1-a)\frac{1}{2} + a\frac{1}{\sqrt{2}} \end{bmatrix}\right) = 2\left((1-a)\frac{1}{2} + a\frac{1}{\sqrt{2}} - 3\right)^4$$

  Αναζητούμε, δηλαδή, το ελάχιστο της συνάρτησης στο [0,1]. Παρατηρούμε πως το ελάχιστο της συνάρτησης είναι το $a_1 = 1$. Επομένως έχουμε πως $x_2 = x_1 + a_1(y_1 - x_1) = y_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

  Παρατηρούμε πως ο υπολογισμός του $y_2 \in S$ τέτοιο ώστε

$$\min_{y\in S} \nabla f(x_2)^T(y - x_2) + \frac{\gamma}{2}\|y - x_2\|_2^2 = \nabla f(x_2)^T(y_2 - x_2) + \frac{\gamma}{2}\|y_2 - x_2\|_2^2$$

  γίνεται με παρόμοιο τρόπο. (Ακριβώς ίδια διαδικασία με Frank-Wolfe).

- *συνέχεια λύσης*: Αναζητούμε $y_2 \in S$ τέτοιο ώστε $\min_{y \in S}\left(\nabla f(x_2)^T(y - x_2) + \frac{\gamma}{2}\|y - x_2\|_2^2\right) = \nabla f(x_2)^T(y_2 - x_2) + \frac{\gamma}{2}\|y_2 - x_2\|_2^2$. Ορίζουμε διάνυσμα $z_2 = x_2 - \frac{1}{\gamma}\nabla f(x_2) := \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} - \frac{1}{1}\begin{bmatrix} -\frac{2}{\sqrt{2}}(3\sqrt{2}-1)^3 \\ -\frac{2}{\sqrt{2}}(3\sqrt{2}-1)^3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}\left(1 + 2(3\sqrt{2}-1)^3\right) \\ \frac{1}{\sqrt{2}}\left(1 + 2(3\sqrt{2}-1)^3\right) \end{bmatrix}$. Επομένως αναζητούμε την προβολή του διανύσματος, $y_2 = P_s z_2$. Εύκολα αποδεικνύεται, από το αποτέλεσμα της προηγούμενης άσκησης έχουμε $y_2 = \frac{1}{\|z_2\|_2} z_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. Επομένως καταλήγουμε στην αναμενόμενη απάντηση $y_2 = x_2$ Εύκολα διαπιστώνουμε πως ο αλγόριθμος τερματίζει καθώς $\delta_2 = \nabla f(x_2)^T(y_2 - x_2) = 0$ και επομένως το x_2 είναι το μοναδικό σημείο ελαχίστου.

- Παρατηρήσεις: Υπολογίσαμε την παράμετρο επιτάχυνσης, αναλυτικά. Στην πράξη, χρησιμοποιούμε προσεγγιστικές διαδικασίες (π.χ τεχνικές γραμμικής αναζητήσεις). Η απλή γεωμετρία του χωρίου βοηθάει καθοριστικά στους υπολογισμούς.

- Το κύριο χαρακτηριστικό της μεθόδου προβεβλημένων κλίσεων είναι η αναζήτηση ένος στοιχείου του συνόλου των περιορισμών (που ορίζει την κατεύθυνση της μεθόδου) το οποίο προκύπτει επιλύοντας ένα πρόβλημα προβολής πάνω στο σύνολο περιορισμών.

- *Άσκηση:* Να χρησιμοποιήσετε την μέθοδο προβεβλημένων κλίσεων, με αρχικό διάνυσμα $x_1 = [0,1/4]^T$, γ=1 για το πρόβλημα $\min_{x=[u,v]^T \in S} u^3 - v^3$ όπου $S = \{[u,v]^T \in \mathbb{R}^2 : u^2 + v^2 \leq 1\}$.

- *Λύση:* Παρατηρούμε πως η συνάρτηση $f(u,v) = u^3 - v^3$ είναι ομαλή, ενώ το σύνολο $S \subset \mathbb{R}^2$ είναι κυρτό και συμπαγές. Επομένως, υπάρχει σημείο ελαχίστου. Παρατηρούμε πως η συνάρτηση δεν είναι κυρτή. Οι υποθέσεις του θεωρήματος σύγκλισης ικανοποιούνται. Υπολογίζουμε $\nabla f(u,v) = [3u^2, -3v^2]^T$. Αναζητούμε $y_1 \in S$ τέτοιο ώστε $\min_{y \in S}\left(\nabla f(x_1)^T(y - x_1) + \frac{\gamma}{2}\|y - x_1\|_2^2\right) = \nabla f(x_1)^T(y_1 - x_1) + \frac{\gamma}{2}\|y_1 - x_1\|_2^2$. Ορίζουμε διάνυσμα $z_1 = x_1 - \frac{1}{\gamma}\nabla f(x_1) := \begin{bmatrix} 0 \\ \frac{1}{4} \end{bmatrix} - \frac{1}{1}\begin{bmatrix} 0 \\ -\frac{3}{16} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{7}{16} \end{bmatrix}$. Επομένως αναζητούμε την προβολή του διανύσματος, $y_1 = P_s z_1$. Εύκολα αποδεικνύεται, από το αποτέλεσμα της προηγούμενης άσκησης έχουμε $y_1 = z_1 = \begin{bmatrix} 0 \\ \frac{7}{16} \end{bmatrix}$.

- <u>*Συνέχεια λύσης*</u>: Υπολογίζουμε $\delta_1 = \nabla f(x_1)^T(y_1 - x_1) = \begin{bmatrix} 0 \\ -\frac{3}{16} \end{bmatrix}^T \begin{bmatrix} 0 \\ \frac{3}{16} \end{bmatrix} < 0.$

  Επομένως ο αλγόριθμος συνεχίζει. Υπολογισμός του $a_1 \in [0,1]$, το οποίο ορίζεται ως: $\min_{a \in [0,1]} f(x_1 + a(y_1 - x_1)) = f(x_1 + a_1(y_1 - x_1))$. Ορίζουμε την συνάρτηση

  $$g(a) = f(x_1 + a(y_1 - x_1)) = f\left(\begin{bmatrix} 0 \\ (1-a)\frac{1}{4} + a\frac{7}{16} \end{bmatrix}\right)$$

  Αναζητούμε, δηλαδή, το ελάχιστο της συνάρτησης στο [0,1]. Παρατηρούμε πως το ελάχιστο της συνάρτησης είναι το $a_1 = 1$. Επομένως έχουμε πως

  $$x_2 = x_1 + a_1(y_1 - x_1) = y_1 = \begin{bmatrix} 0 \\ \frac{7}{16} \end{bmatrix}.$$

  Παρατηρούμε πως ο υπολογισμός του $y_2 \in S$ τέτοιο ώστε $\min_{y \in S} \nabla f(x_2)^T(y - x_2) + \frac{\gamma}{2}\|y - x_k\|_2^2 = \nabla f(x_2)^T(y_2 - x_2) + \frac{\gamma}{2}\|y_2 - x_2\|_2^2$ γίνεται με παρόμοιο τρόπο. (Ακριβώς ίδια διαδικασία με Frank-Wolfe).

- *συνέχεια λύσης*: Αναζητούμε $y_2 \in S$ τέτοιο ώστε
$$\min_{y \in S}\left(\nabla f(x_2)^T(y - x_2) + \frac{\gamma}{2}\|y - x_2\|_2^2 \right) = \nabla f(x_2)^T(y_2 - x_2) + \frac{\gamma}{2}\|y_2 - x_2\|_2^2$$
Ορίζουμε διάνυσμα
$$z_2 = x_2 - \frac{1}{\gamma}\nabla f(x_2) := \begin{bmatrix} 0 \\ \frac{7}{16} \end{bmatrix} - 1 \begin{bmatrix} 0 \\ -3\frac{49}{256} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{259}{256} \end{bmatrix}.$$
Επομένως αναζητούμε την προβολή του διανύσματος, $y_2 = P_s z_2$.
Εύκολα αποδεικνύεται, από το αποτέλεσμα της προηγούμενης άσκησης έχουμε
$y_2 = \frac{1}{\|z_2\|_2} z_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Εύκολα υπολογίζουμε πως $\delta_3 = 0$. Ο αλγόριθμος σταματά και έχουμε βρει υποψήφιο σημείο ελαχίστου το $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

- Βιβλιογραφία:

- Αλ. Μπακόπουλος και Ιων. Χρυσοβέργης, “Εισαγωγή στην Αριθμητική Ανάλυση”, Εκδόσεις Συμμετρία.
- Γ. Ακρίβης και Β. Δουγαλής, “Εισαγωγή στην Αριθμητική Ανάλυση”, Πανεπιστημιακές Εκδόσεις Κρήτης.
- J. Nocedal and S. Wright, “Numerical Optimization”, Springer-Verlag 2006.
- A. Quarteroni, R. Sacco, and F. Saleri, “Numerical Mathematics”, Springer-Verlag 2007.
- E. Polak, “Optimization”, Springer-Verlag 1997