ΜΗ ΓΡΑΜΜΙΚΟΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΧΩΡΙΣ ΠΕΡΙΟΡΙΣΜΟΥΣ

Τα περισσότερα προβλήματα βελτιστοποίησης είναι με περιορισμούς, αλλά οι μέθοδοι επίλυσης χωρίς περιορισμούς έχουν γενικό ενδιαφέρον.

Μέθοδοι που απαιτούν πληροφορία 1ης τάξεως:

- Μέγιστης Κλίσης (Cauchy)
- Fletcher – Reeves
- Newton
- Marquardt

Η ταχύτητα σύγκλισης των περισσότερων μεθόδων βελτιώνεται με κατάλληλη κλιμάκωση των μεταβλητών.

1. Μέγιστη Κλίση – Steepest Descent

   1.1. $X_{q+1} = X_q + aS$
   1.2. $S = -\nabla f$
   1.3. Εύρεση βέλτιστου a^*
   1.4. Έλεγχος $X_{q+1} = X_q - a^* \nabla f\left(X_q\right)$ για optimality
   1.5. Επανάληψη

ΚΡΙΤΗΡΙΑ ΣΥΓΚΛΙΣΗΣ

- $\left|\dfrac{f\left(X_{q+1}\right) - f\left(X_q\right)}{f\left(X_q\right)}\right| \leq \varepsilon$
- $\left|\dfrac{\partial f}{\partial x_i}\right| \leq \varepsilon$, $i = 1, 2, \ldots, n$
- $\left|X_{q+1} - X_q\right| \leq \varepsilon$

ΕΦΑΡΜΟΓΗ 1

<u>Ζητείται:</u> $\min f\left(x_1, x_2\right) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$

Αρχικό σημείο $X_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$

<u>1η επανάληψη</u>

$$\nabla f = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{Bmatrix} = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{Bmatrix}$$

Επομένως, $\nabla f\left(X_1\right) = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$ και $S_1 = -\nabla f\left(X_1\right) = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$

$$X_2 = \begin{Bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{Bmatrix} = X_1 + a_1 S_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + a_1 \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} \Rightarrow \begin{cases} x_1^{(2)} = -a_1 \\ x_2^{(2)} = a_1 \end{cases}$$

Η αντικειμενική συνάρτηση γίνεται:

$$f\left(-a_1, a_1\right) = -a_1 - a_1 + 2\left[a_1\right]^2 - 2\left[a_1\right]^2 + \left[a_1\right]^2 = a_1^{\,2} - 2a_1$$

Για βέλτιστο a_1 θέτουμε $\dfrac{df}{da} = 0 \rightarrow 2a_1 - 2 = 0 \rightarrow \boxed{a_1^* = 1}$

Άρα, $X_2 = X_1 - a_1^* \nabla f\left(X_1\right) = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} - 1 \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$

Επομένως, $\nabla f\left(X_2\right) = \begin{Bmatrix} -1 \\ -1 \end{Bmatrix} \neq \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \rightarrow$ το X_2 όχι βέλτιστο

<u>2η επανάληψη</u>

$$S_2 = -\nabla f\left(X_2\right) = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$$

$$X_3 = \begin{Bmatrix} x_1^{(3)} \\ x_2^{(3)} \end{Bmatrix} = X_2 + a_2 S_2 = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} + a_2 \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \Rightarrow \begin{cases} x_1^{(3)} = -1 + a_2 \\ x_2^{(3)} = 1 + a_2 \end{cases}$$

Η αντικειμενική συνάρτηση γίνεται: $f\left(-1+a_2, 1+a_2\right) = 5a_2^2 - 2a_2 - 1$

Για βέλτιστο a_2 θέτουμε $\dfrac{df}{da} = 0 \rightarrow 10a_2 - 2 = 0 \rightarrow \boxed{a_2^* = 1/5}$

Άρα, $X_3 = X_2 - a_2^* \nabla f\left(X_2\right) = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} + \dfrac{1}{5}\begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -0.8 \\ 1.2 \end{Bmatrix}$

Επομένως, $\nabla f\left(X_3\right) = \begin{Bmatrix} 0.2 \\ -0.2 \end{Bmatrix} \neq \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \rightarrow$ το X_3 όχι βέλτιστο

<u>3η επανάληψη</u>

$$S_3 = -\nabla f\left(X_3\right) = \begin{Bmatrix} -0.2 \\ 0.2 \end{Bmatrix}$$

$$X_4 = \begin{Bmatrix} x_1^{(4)} \\ x_2^{(4)} \end{Bmatrix} = X_3 + a_3 S_3 = \begin{Bmatrix} -0.8 \\ 1.2 \end{Bmatrix} + a_3 \begin{Bmatrix} -0.2 \\ 0.2 \end{Bmatrix} \Rightarrow \begin{cases} x_1^{(4)} = -0.8 - 0.2a_3 \\ x_2^{(4)} = 1.2 + 0.2a_3 \end{cases}$$

Η αντικειμενική συνάρτηση γίνεται: $f\left(-0.8 - 0.2a_3, 1.2 + 0.2a_3\right) = 0.04a_3^2 - 0.08a_3 - 1.20$

Για βέλτιστο a_2 θέτουμε $\dfrac{df}{da} = 0 \rightarrow 0.08a_3 - 0.08 = 0 \rightarrow \boxed{a_3^* = 1}$

Άρα, $X_4 = X_3 - a_3^* \nabla f\left(X_3\right) = \begin{Bmatrix} -0.8 \\ 1.2 \end{Bmatrix} - 1\begin{Bmatrix} -0.2 \\ 0.2 \end{Bmatrix} = \begin{Bmatrix} -0.6 \\ 1.0 \end{Bmatrix}$

Επομένως, $\nabla f\left(X_4\right) = \begin{Bmatrix} 0.6 \\ -0.2 \end{Bmatrix} \neq \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \rightarrow$ το X_4 όχι βέλτιστο

Μετά από επιπλέον επαναλήψεις καταλήγουμε στο βέλτιστο $\boxed{X^* = \begin{Bmatrix} -1.0 \\ 1.5 \end{Bmatrix}}$

ΣΥΖΥΓΕΙΣ ΔΙΕΥΘΥΝΣΕΙΣ (CONJUGATE)

Για κάθε συμμετρικό μητρώο $[A]$ το σύνολο των διανυσμάτων $\{S_i\}$ ορίζει Α-συζυγείς διευθύνσεις όταν:

$$\boxed{S_i^T A S_j = 0} \text{ για } i \neq j \text{ και } i = 1,2,\ldots,n$$

Για $[A] = [I]$ οι διευθύνσεις είναι ορθογωνικές.

Ισχύει: Για μια τετραγωνική συνάρτηση

$$\boxed{Q(X) = C + B^T X + \frac{1}{2} X^T A X}$$

το ελάχιστο βρίσκεται σε λιγότερα από n βήματα ακολουθώντας n συζυγείς διευθύνσεις, ανεξάρτητα από το αρχικό σημείο.

Αν X^* η βέλτιστη λύση, τότε:

$$\boxed{\nabla Q(X^*) = B + AX^* = 0} \quad (1)$$

Για ένα αρχικό σημείο X_1 και n διευθύνσεις S_i, $i = 1,2,\ldots,n$ ισχύει:

$$\boxed{X^* = X_1 + \sum_{i=1}^{n} \beta_i S_i} \quad (2)$$

Που δηλώνει ότι τα διανύσματα S_i, $i = 1,2,\ldots,n$ χρησιμοποιούνται ως βάση.

Αντικαθιστώντας στην (1) προκύπτει:

$$\boxed{B + AX_1 + A\left(\sum_{i=1}^{n} \beta_i S_i\right) = 0} \quad (3)$$

Πολλαπλασιάζοντας με S_j^T λαμβάνουμε:

$$\boxed{S_j^T\left(B + AX_1\right) + S_j^T A\left(\sum_{i=1}^{n} \beta_i S_i\right) = 0} \qquad (4)$$

Αναστρέφοντας την (4) λαμβάνουμε:

$$\boxed{\left(B + AX_1\right)^T S_j + \beta_j S_j^T A S_j = 0} \qquad (5)$$

Καθόσον $\underline{S_i^T A S_j = 0}$ για $i \neq j$

Από την (5) προκύπτει: $$\boxed{\beta_j = -\frac{\left(B + AX_1\right)^T S_j}{S_j^T A S_j}} \qquad (6)$$

Παράλληλα, αν ακολουθηθεί μια επαναληπτική διαδικασία ξεκινώντας από X_1 και ακολουθώντας διαδοχικά τις συζυγείς διευθύνσεις S_i, $i = 1, 2, \ldots, n$, τότε προκύπτουν διαδοχικά τα διανύσματα:

$$\boxed{X_{i+1} = X_i + a_i^* S_i}, \; i = 1, 2, \ldots, n$$

Όπου a_i^* ελαχιστοποιεί την $Q\left(X_i + a_i S_i\right)$, δηλαδή:

$$\boxed{S_i^T \nabla Q\left(X_{i+1}\right) = 0} \qquad (7)$$

Όπου: $$\boxed{\nabla Q\left(X_{i+1}\right) = B + AX_{i+1}} \qquad (8)$$

Οπότε η (7) γράφεται: $$\boxed{S_i^T \left\{B + A\left(X_i + a_i^* S_i\right)\right\} = 0} \qquad (9)$$

Οπότε: $$\boxed{a_i^* = -\frac{\left(B + AX_i\right) S_i^T}{S_i^T A S_i}} \qquad (10)$$

Οπότε: $$\boxed{X_i = X_1 + \sum_{j=1}^{i-1} a_j^* S_j} \qquad (11)$$

Ώστε: $$\boxed{X_i^T A S_i = X_1^T A S_i + \sum_{j=1}^{i-1} a_j^* S_j A S_i = X_1^T A S_i} \qquad (12)$$

Οπότε η σχέση (10) γίνεται: $$\boxed{a_i^* = -\left(B + AX_1\right)^T \frac{S_i}{S_i^T A S_i}} \qquad (13)$$

Η οποία ταυτίζεται με τη σχέση (6) που ορίζει το βέλτιστο διάνυσμα.

Οπότε: $\boxed{\beta_i \equiv a_i^*}$

ΕΦΑΡΜΟΓΗ 2

<u>Ζητείται:</u> $\min f\left(x_1, x_2\right) = 6x_1^2 + 2x_2^2 - 6x_1x_2 - x_1 - 2x_2$

Αν $S_1 = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}$ να βρεθεί η S_2 συζυγής της S_1

$$f\left(X\right) = B^T X + \frac{1}{2} X^T A X = \begin{bmatrix} -1 & -2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 12 & -6 \\ -6 & 4 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}$$

Το μητρώο Hess είναι: $\left[A\right] = \begin{bmatrix} 12 & -6 \\ -6 & 4 \end{bmatrix}$ συμμετρικό και θετικά ορισμένο

Αν $S_2 = \begin{Bmatrix} s_1 \\ s_2 \end{Bmatrix}$ θα πρέπει:

$$S_1^T A S_2 = 0 \Rightarrow \begin{bmatrix} 1 & 2 \end{bmatrix}_{(1\times2)} \begin{bmatrix} 12 & -6 \\ -6 & 4 \end{bmatrix}_{(2\times2)} \begin{Bmatrix} s_1 \\ s_2 \end{Bmatrix} = 0 \Rightarrow \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{Bmatrix} s_1 \\ s_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \Rightarrow \begin{Bmatrix} 0 = 0 \\ 2s_2 = 0 \end{Bmatrix}$$

Άρα s_1 αυθαίρετο, οπότε $\boxed{S_2 = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}}$

ΜΕΘΟΔΟΣ FLETCHER – REEVES

1. Αρχική λύση X_1
2. $S_1 = -\nabla f\left(X_1\right)$ - steepest descent
3. $X_2 = X_1 + a_1^* S_1$ - a_1^* βέλτιστη απόσταση, $i = 2$
4. $\nabla f_i = \nabla f\left(X_i\right)$ και $S_i = -\nabla f_i + \dfrac{\left|\nabla f_i\right|^2}{\left|\nabla f_{i-1}\right|^2} S_{i-1}$
5. $X_{i+1} = X_i + a_i^* S_i$
6. Έλεγχος σύγκλισης → αν δεν έχουμε σύγκλιση τότε πηγαίνουμε στο βήμα 4

ΕΦΑΡΜΟΓΗ 3

<u>Ζητείται:</u> $\min f\left(x_1, x_2\right) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$

Ξεκινώντας με $X_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$

<u>1η επανάληψη</u>

$$\nabla f = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{Bmatrix} = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{Bmatrix}$$

Επομένως, $\nabla f\left(X_1\right) = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$ και $S_1 = -\nabla f\left(X_1\right) = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$

Για να βρεθεί η βέλτιστη απόσταση a_1^* στη διεύθυνση S_1

$f\left(X_1 + a_1 S_1\right) = f\left(-a_1, a_1\right) = a_1^2 - 2a_1$, άρα $\dfrac{df}{da} = 0 \rightarrow \boxed{a_1^* = 1}$

Έτσι: $X_2 = X_1 + a_1^* S_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + 1 \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$

<u>2η επανάληψη:</u>

$$\nabla f\left(X_2\right)=\begin{Bmatrix}-1\\-1\end{Bmatrix}$$

$$\left.\begin{array}{l} S_2=-\nabla f\left(X_2\right)+\dfrac{\left|\nabla f\left(X_2\right)\right|^2}{\left|\nabla f\left(X_1\right)\right|^2}S_1 \\ \left|\nabla f\left(X_2\right)\right|^2=2,\ \left|\nabla f\left(X_1\right)\right|^2=2 \end{array}\right\}\Rightarrow S_2=-\begin{Bmatrix}-1\\-1\end{Bmatrix}+\frac{2}{2}\begin{Bmatrix}-1\\1\end{Bmatrix}=\begin{Bmatrix}0\\2\end{Bmatrix}$$

Για να βρεθεί το a_2^* ελαχιστοποιούμε:

$$f\left(X_2+a_2S_2\right)=f\left(-1,1+2a_2\right)=4a_2^2-2a_2-1,\ \text{άρα}\ \frac{df}{da}=0\rightarrow\boxed{a_2^*=1/4}$$

$$X_3=X_2+a_2^*S_2=\begin{Bmatrix}-1\\1\end{Bmatrix}+\frac{1}{4}\begin{Bmatrix}0\\2\end{Bmatrix}=\begin{Bmatrix}-1\\1.5\end{Bmatrix}$$

<u>3η επανάληψη:</u>

$$\nabla f\left(X_2\right)=\begin{Bmatrix}0\\0\end{Bmatrix}\ \text{optimality}\rightarrow\text{OK, άρα}\ \boxed{X_3=\begin{Bmatrix}-1\\1.5\end{Bmatrix}}\ \text{βέλτιστο}$$

ΜΕΘΟΔΟΣ NEWTON

Μια μη γραμμική συνάρτηση n μεταβλητών μπορεί να αναπτυχθεί σε σειρά Taylor μέχρι όρους $2^{ης}$ τάξης:

$$\boxed{f(X) = f(X_i) + \nabla f(X_i)^T (X - X_i) + \frac{1}{2}(X - X_i)^T [H_i](X - X_i)}$$

$$\text{όπου } H_i = H(X_i) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{|X_i}$$

Για X βέλτιστο ισχύει: $\nabla f(X) = 0$, οπότε: $\nabla f(X) = \nabla f(X_i) + [H_i](X - X_i) = 0$

και άρα $\boxed{X_{i+1} = X_i - [H_i]^{-1} \nabla f_i}$ για $\det[H_i] \neq 0$

ΕΦΑΡΜΟΓΗ 4

Για μια δευτεροβάθμια συνάρτηση η βέλτιστη λύση προκύπτει σε ένα βήμα.

$$\boxed{f(X) = \frac{1}{2} X^T [A] X + B^T X + C}$$

Το ελάχιστο ικανοποιεί την $\boxed{\nabla f(X) = [A] X + B = 0}$, οπότε $X^* = -[A]^{-1} B$

Η επαναληπτική σχέση δίνει:

$$X_{i+1} = X_i - [A]^{-1} \nabla f_i = X_i - [A]^{-1}([A] X_i + B) = X_i - [A]^{-1}[A] X_i - [A]^{-1} B = -[A]^{-1} B$$

Και άρα για X_i αρχική τιμή: $\boxed{X_{i+1} = X^* = -[A]^{-1} B}$

*Η μέθοδος απαιτεί δευτεροβάθμια πληροφορία δύσκολη και απαιτητική υπολογιστικά.

ΜΕΘΟΔΟΣ MARQUARDT

Η μέθοδος μέγιστης κλίσης είναι αποδοτική για X_i μακριά της βέλτιστης λύσης, ενώ η μέθοδος Newton πολύ αποτελεσματική κοντά στη βέλτιστη λύση. Η μέθοδος Marquardt επιχειρεί να συνδυάσει τα πλεονεκτήματα και των δύο μεθόδων. Τούτο επιτυγχάνεται τροποποιώντας το μητρώο Hess ως εξής:

$$\boxed{\left[\tilde{H}\left(X_i\right)\right]=\left[H\left(X_i\right)\right]+a_i\left[I\right]}$$

Όπου $\left[I\right]$ το μοναδιαίο τετραγωνικό μητρώο και $a_i > 0$

Η αύξηση των διαγώνιων όρων του μητρώου Hess εξασφαλίζει το θετικά ορισμένο του τροποποιημένου μητρώου Hess.

Για $a_i > 10^4$:

$$\boxed{\left[\tilde{H}\left(X_i\right)\right]^{-1}=\left[\left[H\left(X_i\right)\right]+a_i\left[I\right]\right]^{-1}\cong\frac{1}{a_i}\left[I\right]}$$

ΕΛΕΓΧΟΣ ΣΥΝΘΗΚΩΝ KUHN – TUCKER

Για ένα μη γραμμικό πρόβλημα n μεταβλητών X_i, $i = 1,2,\ldots,n$ ισχύει:

$$\boxed{\sum_{j\in J} \lambda_j \frac{\partial g_j}{\partial x_i} = -\frac{\partial f}{\partial x_i}},\ i = 1,2,\ldots,n$$

Όπου J είναι το σύνολο των ενεργών περιορισμών ανισότητας. Η παραπάνω σχέση γράφεται σε μητρωική μορφή:

$$\boxed{\underset{n\times p}{[G]}\underset{p\times 1}{\{\lambda\}} = \underset{n\times 1}{\{F\}}}$$

Όπου $[G] = \underset{n\times p}{\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_2}{\partial x_1} & \cdots & \frac{\partial g_p}{\partial x_1} \\ \frac{\partial g_1}{\partial x_2} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_p}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_n} & \frac{\partial g_2}{\partial x_n} & \cdots & \frac{\partial g_p}{\partial x_n} \end{bmatrix}_X}$, $\{\lambda\} = \begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_p \end{Bmatrix}$ και $\{F\} = \begin{Bmatrix} -\frac{\partial f}{\partial x_1} \\ -\frac{\partial f}{\partial x_2} \\ \vdots \\ -\frac{\partial f}{\partial x_n} \end{Bmatrix}_X$

υπολογισμένα στο υπόψη σημείο X

Οι συντελεστές Lagrange προκύπτουν: $\boxed{\underset{p\times 1}{\{\lambda\}} = \underset{p\times p}{\left[\underset{p\times n}{G^T}\ \underset{n\times p}{G}\right]^{-1}} \underset{p\times n}{G^T} \underset{n\times 1}{\{F\}}}$

Αν όλα τα $\lambda > 0$ τότε X - optimum

Η δυσκολία είναι να προσδιοριστεί το σύνολο των ενεργών περιορισμών. Συνήθως:

$$\left|g_j(x)\right| \le \varepsilon \text{ για } 10^{-2} \le \varepsilon \le 10^{-6}$$

*Όλοι οι περιορισμοί πρέπει να είναι κανονικοποιημένοι.