ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Εργαστήριο Θερμικών Στροβιλομηχανών
Μονάδα Παράλληλης Υπολογιστικής Ρευστοδυναμικής & Βελτιστοποίησης

ΜΕΘΟΔΟΙ ΑΕΡΟΔΥΝΑΜΙΚΗΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

(7ο Εξάμηνο Σχολής Μηχ.Μηχ. ΕΜΠ)

ΑΙΤΙΟΚΡΑΤΙΚΗ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Κυριάκος Χ. Γιαννάκογλου
Καθηγητής ΕΜΠ

kgianna@central.ntua.gr

http://velos0.ltt.mech.ntua.gr/research/

Γιατί ασχολούμαστε με τετραγωνικές μορφές;

Ορισμός 3.1 *Ορίζουμε ως* **τετραγωνική μορφή** *(quadratic form) ενός διανύσματος* $\overrightarrow{x}$ *κάθε βαθμωτή συνάρτηση της μορφής*

$$F(\overrightarrow{x}) = \frac{1}{2}\overrightarrow{x}^T A\overrightarrow{x} - \overrightarrow{b}^T\overrightarrow{x} + c \tag{3.7}$$

όπου A ένα τετραγωνικό μητρώο και c μια βαθμωτή σταθερά.

Θεώρημα 3.1 *Αν το A είναι συμμετρικό και θετικά ορισμένο μητρώο, τότε το ελάχιστο της $F(\overrightarrow{x})$ (σχέση 3.7) ταυτίζεται με τη λύση του γραμμικού συστήματος*

$$A\overrightarrow{x} = \overrightarrow{b}$$

όπου A ένα τετραγωνικό μητρώο και c μια βαθμωτή σταθερά.

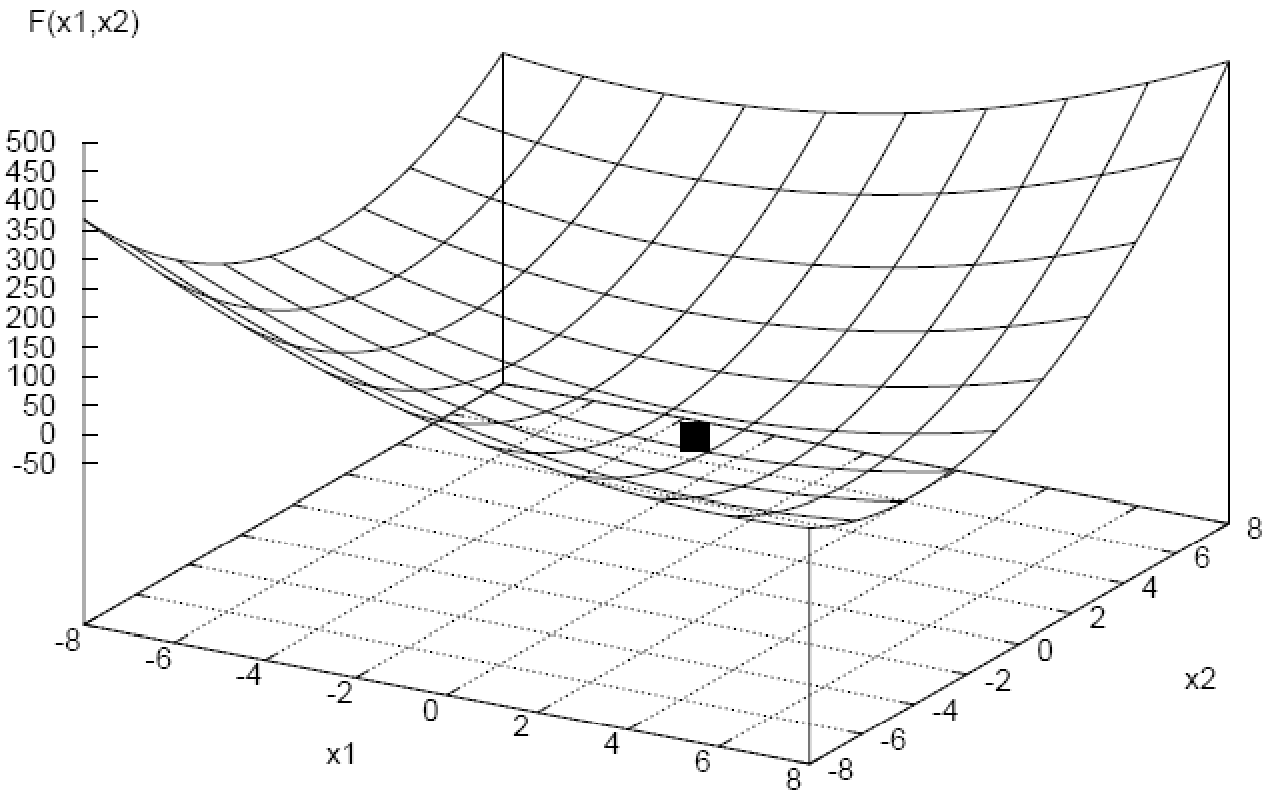
$$\nabla F(\overrightarrow{x}) = \frac{1}{2}A^T\overrightarrow{x} + \frac{1}{2}A\overrightarrow{x} - \overrightarrow{b}$$

$$\nabla F(\overrightarrow{x}) = A\overrightarrow{x} - \overrightarrow{b}$$

Γιατί ασχολούμαστε με τετραγωνικές μορφές;

$$
\begin{aligned}
F(\overrightarrow{x}^* + \overrightarrow{e}) &= \frac{1}{2}(\overrightarrow{x}^* + \overrightarrow{e})^T A(\overrightarrow{x}^* + \overrightarrow{e}) - \overrightarrow{b}^T(\overrightarrow{x}^* + \overrightarrow{e}) + c \\
&= \frac{1}{2}\overrightarrow{x}^{*T} A\overrightarrow{x}^* + \overrightarrow{e}^T A\overrightarrow{x}^* + \frac{1}{2}\overrightarrow{e}^T A\overrightarrow{e} - \overrightarrow{b}^T\overrightarrow{x}^* - \overrightarrow{b}^T\overrightarrow{e} + c \\
&= \frac{1}{2}\overrightarrow{x}^{*T} A\overrightarrow{x}^* - \overrightarrow{b}^T\overrightarrow{x}^* + c + \overrightarrow{e}^T\overrightarrow{b} + \frac{1}{2}\overrightarrow{e}^T A\overrightarrow{e} - \overrightarrow{b}^T\overrightarrow{e} \\
&= F(\overrightarrow{x}^*) + \frac{1}{2}\overrightarrow{e}^T A\overrightarrow{e}
\end{aligned}
$$

Γραφική παράσταση τετραγωνικής Συνάρτησης



Απότομη Κάθοδος για Γραμμικά Προβλήματα

Μέθ. Απότομης Καθόδου

$$\vec{x}^{\,n+1} = \vec{x}^{\,n} + \eta \, \vec{p}^{\,n}$$
$$\vec{p}^{\,n} = -\nabla F(\vec{x}^{\,n})$$

Για την ελαχιστοποίηση της

$$F(\vec{x}) = \frac{1}{2}\vec{x}^{\,T} A \vec{x} - \vec{b}^{\,T}\vec{x} + c$$

Είναι ως να λύνεται η

$$A\,\vec{x} = \vec{b}$$

Με υπόλοιπο της εξίσωσης το

$$\vec{r}^{\,n} = \vec{b} - A\vec{x}^{\,n}$$

Άρα θα είναι

$$\vec{r}^{\,n} = \vec{p}^{\,n} = -\nabla F(\vec{x}^{\,n}) = \vec{b} - A\vec{x}^{\,n}$$

$$\vec{x}^{\,n+1} = \vec{x}^{\,n} + \eta^{n}\, \vec{r}^{\,n}$$

Απότομη Κάθοδος για Γραμμικά Προβλήματα

Ποιο το βέλτιστο βήμα η ;

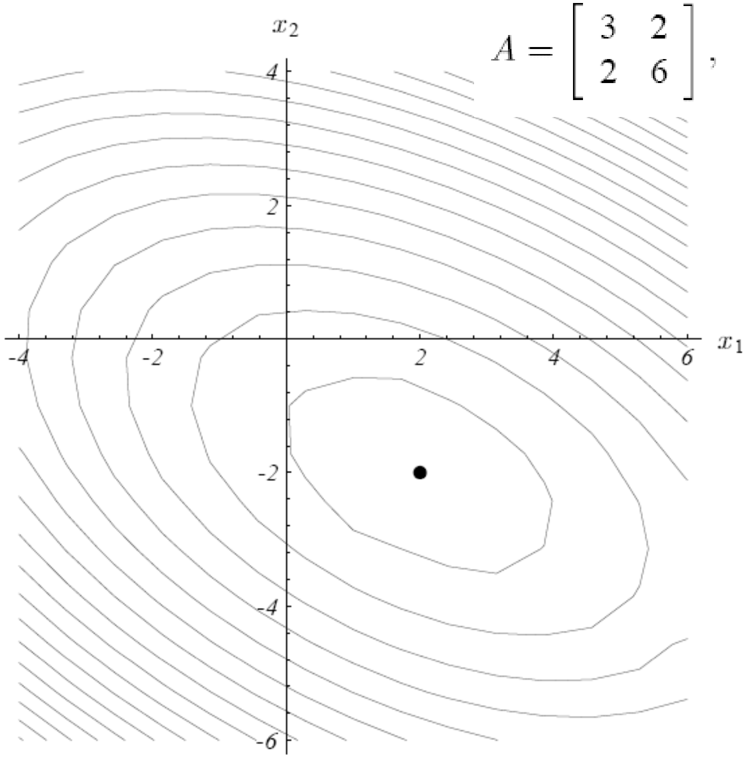
$$\frac{d}{d\eta}F(\overrightarrow{x}^{n+1}) \;=\; \nabla F(\overrightarrow{x}^{n+1})^T \frac{d}{d\eta}\overrightarrow{x}^{n+1} \;=\; 0$$

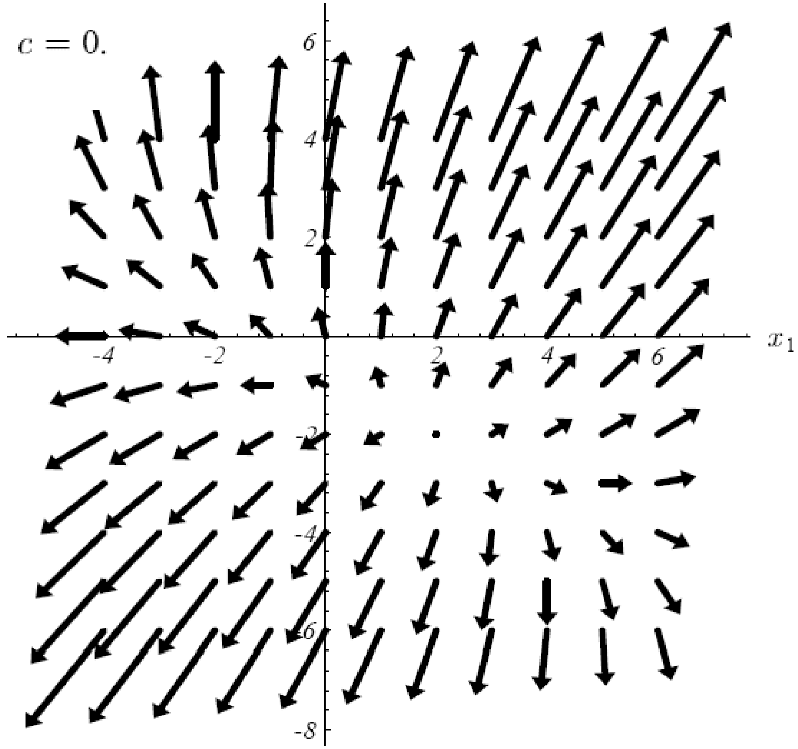
ή

$$\nabla F(\overrightarrow{x}^{n+1})^T \overrightarrow{r}^{n} \;=\; -\overrightarrow{r}^{n+1^T} \overrightarrow{r}^{n} \;=\; 0$$

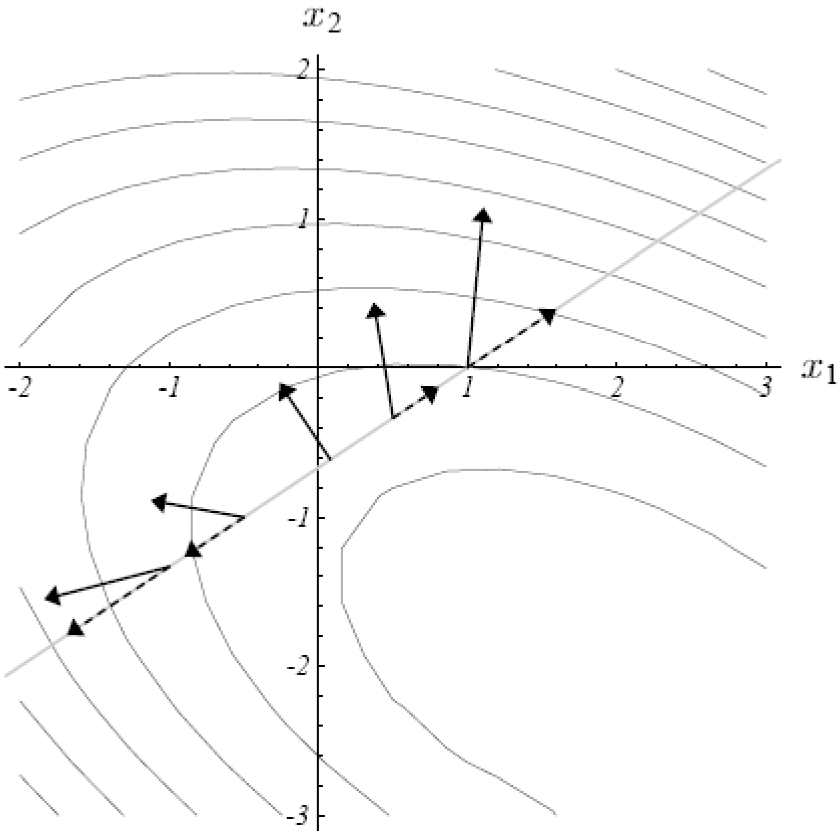
Κατανόηση του τρόπου που προχωρά η Απότομη Κάθοδος:

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}, \qquad b = \begin{bmatrix} 2 \\ -8 \end{bmatrix}, \qquad c = 0.$$





Έστω ότι ξεκινώ από το (-2,-2):



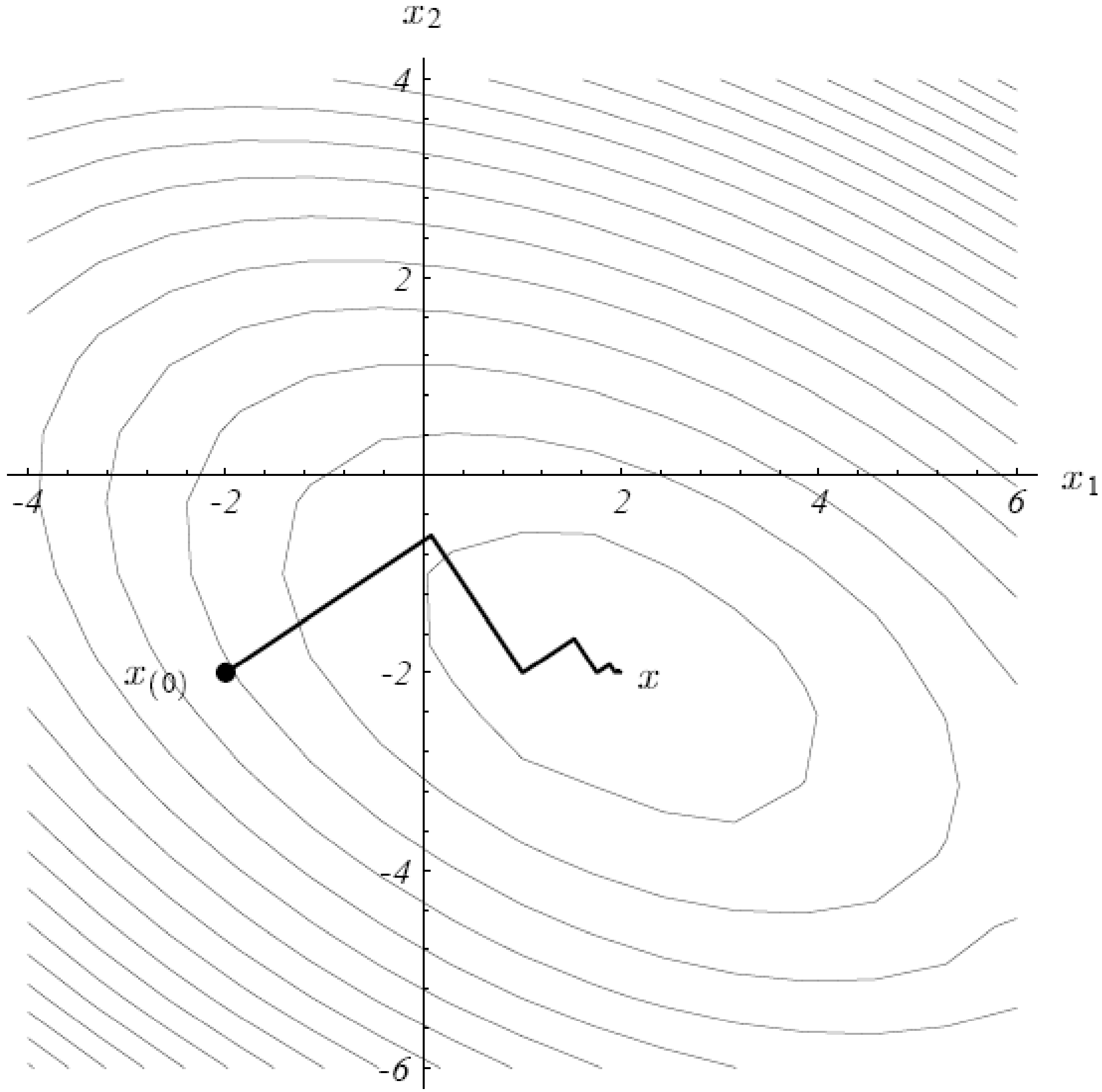
Παράδειγμα από το:

An Introduction to
the Conjugate Gradient Method
Without the Agonizing Pain
Edition $1\frac{1}{4}$

Jonathan Richard Shewchuk
August 4, 1994

Απότομη Κάθοδος για Γραμμικά Προβλήματα

Έστω ότι ξεκινώ από το (-2,-2). Πως καταλήγω στο (2,-2):



Απότομη Κάθοδος για Γραμμικά Προβλήματα

Υπολογισμός του (βέλτιστου) βήματος η:

$$
\begin{aligned}
\overrightarrow{r}^{n+1^T}\overrightarrow{r}^n &= 0\\
(\overrightarrow{b} - A\overrightarrow{x}^{n+1})^T\overrightarrow{r}^n &= 0\\
(\overrightarrow{b} - A(\overrightarrow{x}^n + \eta^n\overrightarrow{r}^n))^T\overrightarrow{r}^n &= 0\\
(\overrightarrow{b} - A\overrightarrow{x}^n)^T\overrightarrow{r}^n - \eta^n(A\overrightarrow{r}^n)^T\overrightarrow{r}^n &= 0\\
(\overrightarrow{b} - A\overrightarrow{x}^n)^T\overrightarrow{r}^n &= \eta^n(A\overrightarrow{r}^n)^T\overrightarrow{r}^n\\
\overrightarrow{r}^{n^T}\overrightarrow{r}^n &= \eta^n\overrightarrow{r}^{n^T}A\overrightarrow{r}^n
\end{aligned}
$$

Άρα, τελικός τύπος για το η:

$$
\eta^n = \frac{\overrightarrow{r}^{n^T}\overrightarrow{r}^n}{\overrightarrow{r}^{n^T}A\overrightarrow{r}^n}
$$

Απότομη Κάθοδος για Γραμμικά Προβλήματα

Αλγόριθμος:

Βήμα 0: Επιλογή αρχικής λύσης $\vec{x}^0$. Δείκτης $n=0$.

Βήμα 1: Υπολογισμός υπολοίπου $\vec{r}^n = \vec{b} - A\vec{x}^n$.

Βήμα 2: Υπολογισμός μεγέθους βήματος $\eta^n = \frac{\vec{r}^{n^T}\vec{r}^n}{\vec{r}^{n^T}A\vec{r}^n}$,

Βήμα 3: Ανανέωση λύσης $\vec{x}^{n+1} = \vec{x}^n + \eta^n\vec{r}^n$.

Βήμα 4: Ανανέωση τιμής δείκτη $n \leftarrow n+1$. Επιστροφή στο βήμα 1, μέχρι σύγκλισης.

Σχόλια περί υπολογιστικού κόστους …

“*matrix-vector products*”

Οικονομία:

$$\begin{aligned}
\vec{x}^{n+1} &= \vec{x}^n + \eta^n\,\vec{r}^n \\
\vec{b} - A\vec{x}^{n+1} &= \vec{b} - A\vec{x}^n - \eta^n\,A\vec{r}^n \\
\vec{r}^{n+1} &= \vec{r}^n - \eta^n\,A\vec{r}^n
\end{aligned}$$

Νέος Αλγόριθμος:

Βήμα 0: Επιλογή αρχικής λύσης $\overrightarrow{x}^0$. Δείκτης $n = 0$. Υπολογισμός αρχικού υπολοίπου $\overrightarrow{r}^n = \overrightarrow{b} - A\overrightarrow{x}^n$.

Βήμα 1: Υπολογισμός μεγέθους βήματος $\eta^n = \frac{\overrightarrow{r}^{nT}\overrightarrow{r}^n}{\overrightarrow{r}^{nT}A\overrightarrow{r}^n}$, με αποθήκευση του διανύσματος $A\overrightarrow{r}^n$ για επόμενη χρήση.

Βήμα 2: Ανανέωση λύσης $\overrightarrow{x}^{n+1} = \overrightarrow{x}^n + \eta^n\overrightarrow{r}^n$.

Βήμα 3: Ανανέωση υπολοίπου $\overrightarrow{r}^{n+1} = \overrightarrow{r}^n - \eta^n\, A\overrightarrow{r}^n$, χρησιμοποιώντας το ήδη υπολογισμένο διάνυσμα $A\overrightarrow{r}^n$. Ανανέωση τιμής δείκτη $n \leftarrow n+1$ και επιστροφή στο βήμα 1, μέχρι σύγκλισης.

Τι πρόβλημα/μειονέκτημα έχει αυτός;

Η Έννοια της Συζυγίας

Είναι η αφετηρία για τη θεμελίωση της Μεθόδου των Συζυγών Κατευθύνσεων!

Ιδέα: Να μην ξαναχρησιμοποιείται κατεύθυνση που ήδη σαρώθηκε!

Ορισμός κατευθύνσεων (το πολύ N) $\quad \vec{p}^{\,0}, \vec{p}^{\,1}, \vec{p}^{\,2}, \ldots, \vec{p}^{\,N-1}$

τέτοιων ώστε:

$$\vec{p}^{\,iT} \vec{p}^{\,j} = 0 \,, \qquad i \neq j$$

Στο γραμμικό πρόβλημα, με την απότομη κάθοδο, το σφάλμα (error):

$$\vec{e}^{\,n+1} = \vec{e}^{\,n} + \eta^n \, \vec{p}^{\,n}$$

Απαίτηση:

$$\vec{p}^{\,nT} \, \vec{e}^{\,n+1} = 0$$

Ή

$$\vec{p}^{\,nT} \, \vec{e}^{\,n} + \eta^n \vec{p}^{\,nT} \vec{p}^{\,n} = 0$$

Κι αυτό;;;

Άρα:

$$\eta^n = - \frac{\vec{p}^{\,nT} \vec{e}^{\,n}}{\vec{p}^{\,nT} \vec{p}^{\,n}}$$

Μέθοδος Συζυγών Κατευθύνσεων (Conjugate Directions)

Συζυγία ή Α-ορθογώνια διανύσματα:

δηλαδή απαιτώ: $$\vec{p}^{iT} A \vec{p}^{j} = 0 \,, \qquad i \neq j$$

Ισχύει:

$$\begin{aligned} \frac{dF(\vec{x}^{n+1})}{d\eta^n} &= 0 \\ \nabla F(\vec{x}^{n+1}) \frac{d\vec{x}^{n+1}}{d\eta^n} &= 0 \\ -\vec{r}^{n+1^T} \vec{p}^{n} &= 0 \end{aligned}$$

Όμως:

$$\vec{r}^{n+1} = \vec{b} - A\vec{x}^{n+1} = \vec{b} - A(\vec{x}^{*} + \vec{e}^{n+1})$$

$$\vec{r}^{n+1} = -A\vec{e}^{n+1}$$

Άρα:

$$\vec{p}^{nT} A \vec{e}^{n+1} = 0$$

Μέθοδος Συζυγών Κατευθύνσεων (Conjugate Directions)

Με τη νέα απαίτηση, το βέλτιστο βήμα η είναι:

$$\begin{aligned} \overrightarrow{p}^{nT} A \overrightarrow{e}^{n+1} &= 0 \\ \overrightarrow{p}^{nT} A(\overrightarrow{e}^{n} + \eta^n \overrightarrow{p}^{n}) &= 0 \\ \overrightarrow{p}^{nT} A \overrightarrow{e}^{n} + \eta^n \overrightarrow{p}^{nT} A \overrightarrow{p}^{n} &= 0 \\ -\overrightarrow{p}^{nT} \overrightarrow{r}^{n} + \eta^n \overrightarrow{p}^{nT} A \overrightarrow{p}^{n} &= 0 \end{aligned}$$

Άρα, νέα χρήσιμη έκφραση η:

$$\eta^n = \frac{\overrightarrow{p}^{nT} \overrightarrow{r}^{n}}{\overrightarrow{p}^{nT} A \overrightarrow{p}^{n}}$$

Μια τέτοια μέθοδος συγκλίνει <u>το πολύ σε Ν βήματα</u>. Η απόδειξη υπάρχει στο βιβλίο, αλλά μπορείτε εύκολα να την καταλάβετε με βάση την ιδέα που εφαρμόστηκε.

Τι δεν είπαμε: Πως θα βρεθούν τα $\overrightarrow{p}^{0}, \overrightarrow{p}^{1}, \overrightarrow{p}^{2}, \ldots, \overrightarrow{p}^{N-1}$

Συζυγής Διαδικασία των Gram-Schmidt

$$\vec{p}^{\,n} = \vec{u}^{\,n} + \sum_{i=0}^{n-1} \beta_{ni} \vec{p}^{\,i}$$

όπου $\vec{u}^{\,0}, \vec{u}^{\,1}, \vec{u}^{\,2}, \cdots, \vec{u}^{\,N-1}$ **είναι N γραμμικά ανεξάρτητα διανύσματα.**

Λ.χ. οι συντελεστές για N=4:

$$\begin{bmatrix} - & - & - & - \\ \beta_{10} & - & - & - \\ \beta_{20} & \beta_{21} & - & - \\ \beta_{30} & \beta_{31} & \beta_{32} & - \end{bmatrix}$$

Απαίτηση για Α-ορθογωνιότητα:

$$\left.\begin{aligned} \vec{p}^{\,iT} A \vec{p}^{\,j} &= 0 \\ \left(\vec{u}^{\,i} + \sum_{m=0}^{i-1} \beta_{im} \vec{p}^{\,m}\right)^T A \vec{p}^{\,j} &= 0 \\ \vec{u}^{\,iT} A \vec{p}^{\,j} + \sum_{m=0}^{i-1} \beta_{im} \vec{p}^{\,mT} A \vec{p}^{\,j} &= 0 \\ \vec{u}^{\,iT} A \vec{p}^{\,j} + \beta_{ij} \vec{p}^{\,jT} A \vec{p}^{\,j} &= 0 \end{aligned}\right\} \quad \beta_{ij} = -\frac{\vec{u}^{\,iT} A \vec{p}^{\,j}}{\vec{p}^{\,jT} A \vec{p}^{\,j}} \quad , \; i > j$$

Μέθοδος Συζυγών Κλίσεων (Conjugate Gradient, CG)

Χρησιμοποιεί τα υπόλοιπα r^i στη θέση των u^i στην προηγούμενη μέθοδο. Άρα:

$$\beta_{ij} = -\frac{\vec{r}^{iT} A \vec{p}^{j}}{\vec{p}^{jT} A \vec{p}^{j}} \qquad , i > j$$

Στη Gram-Schmidt μπορεί να δειχτεί (βλ. βιβλίο) ότι:

$$\vec{r}^{n+1} = \vec{r}^{n} - \eta^{n} A \vec{p}^{n}$$

Άρα:

$$\eta^{j} \vec{r}^{iT} A \vec{p}^{j} = \vec{r}^{iT} \vec{r}^{j} - \vec{r}^{iT} \vec{r}^{j+1}$$

Καταλήγουμε ότι:

$$\vec{r}^{iT} A \vec{p}^{j} = \begin{cases} \frac{1}{\eta^j} \vec{r}^{jT} \vec{r}^{j} & , i = j \\ -\frac{1}{\eta^j} \vec{r}^{j+1T} \vec{r}^{j+1} & , i = j+1 \\ 0 & , otherwise \end{cases}$$

Δηλαδή:

$$\beta_{ij} = \begin{cases} \frac{1}{\eta^{i-1}} \frac{\vec{r}^{iT} \vec{r}^{i}}{\vec{p}^{i-1T} A \vec{p}^{i-1}} & , i = j+1 \\ \\ 0 & , otherwise \end{cases}$$

Μέθοδος Συζυγών Κλίσεων (Conjugate Gradient, CG)

Αν:

$$\beta^i \quad \longleftrightarrow \quad \beta_{i,i-1}$$

Τότε:

$$\beta^i = \frac{1}{\eta^{i-1}} \frac{\overrightarrow{r}^{iT} \overrightarrow{r}^{i}}{\overrightarrow{p}^{i-1^T} A \overrightarrow{p}^{i-1}}$$

$$\begin{bmatrix} - & - & - & - \\ \beta^1 & - & - & - \\ - & \beta^2 & - & - \\ - & - & \beta^3 & - \end{bmatrix}$$

Και αφού:

$$\eta^{i-1} = \frac{\overrightarrow{p}^{i-1^T} \overrightarrow{r}^{i-1}}{\overrightarrow{p}^{i-1^T} A \overrightarrow{p}^{i-1}} \Rightarrow \eta^{i-1} \overrightarrow{p}^{i-1^T} A \overrightarrow{p}^{i-1} = \overrightarrow{p}^{i-1^T} \overrightarrow{r}^{i-1}$$

$$\beta^i = \frac{\overrightarrow{r}^{iT} \overrightarrow{r}^{i}}{\overrightarrow{p}^{i-1^T} \overrightarrow{r}^{i-1}}$$

Ή αφού:

$$\overrightarrow{p}^{iT} \overrightarrow{r}^{i} = \overrightarrow{r}^{iT} \overrightarrow{r}^{i} \Rightarrow \overrightarrow{p}^{i-1^T} \overrightarrow{r}^{i-1} = \overrightarrow{r}^{i-1^T} \overrightarrow{r}^{i-1}$$

$$\beta^i = \frac{\overrightarrow{r}^{iT} \overrightarrow{r}^{i}}{\overrightarrow{r}^{i-1^T} \overrightarrow{r}^{i-1}}$$

Μέθοδος Συζυγών Κλίσεων (Conjugate Gradient, CG)

Βήμα 0: Επιλογή αρχικής λύσης $\vec{x}^0$. Υπολογισμός αρχικού υπολοίπου $\vec{r}^0 = \vec{b} - A\vec{x}^0$. Ορισμός αρχικής κατεύθυνσης ανίχνευσης $\vec{p}^0 = \vec{r}^0$. Δείκτης $n = 0$.

Βήμα 1: Υπολογισμός μεγέθους βήματος $\eta^n = \frac{\vec{r}^{nT}\vec{r}^n}{\vec{p}^{nT}A\vec{p}^n}$, (σχέση 3.47, σε συνδυασμό με τη σχέση 3.69).

Βήμα 2: Ανανέωση λύσης $\vec{x}^{n+1} = \vec{x}^n + \eta^n\vec{p}^n$.

Βήμα 3: Ανανέωση υπολοίπου $\vec{r}^{n+1} = \vec{r}^n - \eta^n A\vec{p}^n$, (σχέση 3.61).

Βήμα 4: Υπολογισμός συντελεστή $\beta^{n+1} = \frac{\vec{r}^{n+1T}\vec{r}^{n+1}}{\vec{r}^{nT}\vec{r}^n}$, (σχέση 3.70).

Βήμα 5: Ανανέωση διανύσματος κατεύθυνσης ανίχνευσης $\vec{p}^{n+1} = \vec{r}^{n+1} + \beta^{n+1}\vec{p}^n$,(προκύπτει από τη σχέση 3.54, για δείκτη $n+1$ αντί του n, με το $\vec{r}^{n+1}$ αντί του $\vec{u}^{n+1}$ λόγω της γνωστής παραδοχής και κρατώντας τη μοναδική μη-μηδενική τιμή του συντελεστή β από όλους τους όρους της άθροισης).

Βήμα 6: Ανανέωση τιμής δείκτη $n \leftarrow n+1$. Επιστροφή στο βήμα 1, μέχρι σύγκλισης.

Με:

$$\overrightarrow{r}^{n} = -\nabla F(\overrightarrow{x}^{n})$$

Βήμα 0: Επιλογή αρχικής λύσης $\overrightarrow{x}^0$. Υπολογισμός $F(\overrightarrow{x}^0)$. Υπολογισμός αρχικού υπολοίπου $\overrightarrow{r}^0 = -\nabla F(\overrightarrow{x}^0)$, (σχέση 3.71). Ορισμός αρχικής κατεύθυνσης ανίχνευσης $\overrightarrow{p}^0 = \overrightarrow{r}^0$. Δείκτης $n = 0$.

Βήμα 1: Αν $\overrightarrow{r}^n \neq 0$, υπολογισμός μεγέθους βήματος η^n με (ακριβή ή έστω προσεγγιστική) ανίχνευση κατά γραμμή, στην κατεύθυνση $\overrightarrow{p}^n$, της σχέσης $\min_{\eta^n>0} F(\overrightarrow{x}^n + \eta^n \overrightarrow{p}^n)$.

Βήμα 2: Ανανέωση λύσης $\overrightarrow{x}^{n+1} = \overrightarrow{x}^n + \eta^n \overrightarrow{p}^n$.

Βήμα 3: Υπολογισμός $\nabla F(\overrightarrow{x}^{n+1})$.

Βήμα 4: Υπολογισμός συντελεστή β^{n+1} από τη σχέση

$$\beta_{FR}^{n+1} = \frac{\nabla F(\overrightarrow{x}^{n+1})^T \nabla F(\overrightarrow{x}^{n+1})}{\nabla F(\overrightarrow{x}^{n})^T \nabla F(\overrightarrow{x}^{n})} \qquad (3.72)$$

Βήμα 5: Ανανέωση διανύσματος κατεύθυνσης ανίχνευσης $\overrightarrow{p}^{n+1} = -\nabla F(\overrightarrow{x}^{n+1}) + \beta_{FR}^{n+1} \overrightarrow{p}^n$.

Βήμα 6: Ανανέωση τιμής δείκτη $n \leftarrow n+1$. Επιστροφή στο βήμα 1, μέχρι σύγκλισης.

Υπολογισμός Εσσιανής (Hessian Matrix Computation)

Newton Methods :

$$b_i^{n+1} = b_i^n + db_i$$

$$\frac{d^2F}{db_i db_j} db_j = -\frac{dF}{db_i}$$

Hessian Matrix Computation using the DD-DD method:

(Think "Discrete", program Continuous Adjoint)

$$\frac{dF}{db_i} = \frac{\partial F}{\partial b_i} + \frac{\partial F}{\partial U_k}\frac{dU_k}{db_i}$$

k=1,…,N design variables

$$\frac{d^2F}{db_i db_j} = \frac{\partial^2 F}{\partial b_i \partial b_j} + \frac{\partial^2 F}{\partial b_i \partial U_k}\frac{dU_k}{db_j} + \frac{\partial^2 F}{\partial U_k \partial b_j}\frac{dU_k}{db_i} + \frac{\partial^2 F}{\partial U_k \partial U_m}\frac{dU_k}{db_i}\frac{dU_m}{db_j} + \frac{\partial F}{\partial U_k}\frac{d^2 U_k}{db_i db_j}$$

$$\frac{dR_m}{db_i} = \frac{\partial R_m}{\partial b_i} + \frac{\partial R_m}{\partial U_k}\frac{dU_k}{db_i} = 0$$

$$\frac{d^2R_n}{db_i db_j} = \frac{\partial^2 R_n}{\partial b_i \partial b_j} + \frac{\partial^2 R_n}{\partial b_i \partial U_k}\frac{dU_k}{db_j} + \frac{\partial^2 R_n}{\partial U_k \partial b_j}\frac{dU_k}{db_i} + \frac{\partial^2 R_n}{\partial U_k \partial U_m}\frac{dU_k}{db_i}\frac{dU_m}{db_j} + \frac{\partial R_n}{\partial U_k}\frac{d^2 U_k}{db_i db_j} = 0$$

► **The cost for computing the Hessian via the DD-DD approach scales with N².**

Computation of the Hessian Matrix, via DD-AV

$$\frac{dR_m}{db_i} = \frac{\partial R_m}{\partial b_i} + \frac{\partial R_m}{\partial U_k}\frac{dU_k}{db_i} = 0 \quad \Rightarrow \quad \boxed{\frac{dU_k}{db_i}} \qquad \boxed{N}$$

System solutions (EFS)

$$\frac{\partial F}{\partial U_k} + \hat{\Psi}_n \frac{\partial R_n}{\partial U_k} = 0 \quad \Rightarrow \quad \boxed{\hat{\Psi}_n} \qquad \boxed{1}$$

EFS

The Adjoint equation is the same with that obtained for the Gradient !!!

$$\frac{d^2\hat{F}}{db_i db_j} = \frac{\partial^2 F}{\partial b_i \partial b_j} + \hat{\Psi}_n \frac{\partial^2 R_n}{\partial b_i \partial b_j} + \frac{\partial^2 F}{\partial U_k \partial U_m}\frac{dU_k}{db_i}\frac{dU_m}{db_j} + \hat{\Psi}_n \frac{\partial^2 R_n}{\partial U_k \partial U_m}\frac{dU_k}{db_i}\frac{dU_m}{db_j}$$
$$+ \frac{\partial^2 F}{\partial b_i \partial U_k}\frac{dU_k}{db_j} + \hat{\Psi}_n \frac{\partial^2 R_n}{\partial b_i \partial U_k}\frac{dU_k}{db_j} + \frac{\partial^2 F}{\partial U_k \partial b_j}\frac{dU_k}{db_i} + \hat{\Psi}_n \frac{\partial^2 R_n}{\partial U_k \partial b_j}\frac{dU_k}{db_i}$$
$$+ \cancel{\left(\frac{\partial F}{\partial U_k} + \hat{\Psi}_n \frac{\partial R_n}{\partial U_k}\right)} \frac{d^2 U_k}{db_i db_j}$$

$$\frac{d^2 F^\lambda}{db_i db_j} db_j^\lambda = -\frac{dF^\lambda}{db_i}$$

► **The cost per Newton cycle is N+1+1=N+2 EFS! Scales with N, not N^2.**

► **DD-AV is the most efficient approach (among DD-DD, AV-DD, AV-AV)!**

Truncated Newton Methods

Why?

- ► Create more efficient Hessian-based optimization schemes.
- ► Desirable cost per cycle **<<N EFS**, without damaging accuracy.

Inspired by:
The **Conjugate Gradient (CG)** method for solving systems of linear equations

$$\boxed{Ax = q}$$

requires only matrix-vector products.

$k \leftarrow 0$

$x \leftarrow \text{init}()$

$r^0 \leftarrow \boxed{Ax} - q;\ p \leftarrow -r^0$

while $r^k \neq 0$ **and** $k \leq M_{CG}$ **do**

$\quad \eta \leftarrow \frac{(r^k)^T r^k}{p^T \boxed{Ap}}$

$\quad x \leftarrow x + \eta p$

$\quad r^{k+1} \leftarrow r^k + \eta \boxed{Ap}$

$\quad \beta \leftarrow \frac{(r^{k+1})^T r^{k+1}}{(r^k)^T r^k}$

$\quad p \leftarrow -r_{k+1} + \beta p$

$\quad k \leftarrow k + 1$

end while

Truncated Newton Methods

$$b_i^{n+1} = b_i^n + db_i$$

$$\frac{d^2F}{db_i db_j} db_j = -\frac{dF}{db_i}$$

$k \leftarrow 0$
$b_j \leftarrow \text{init}()$
while $k \leq k_{max}$ **do**
  $U_n \leftarrow$ Flow Equations [1 EFS] ★
  $\Psi_n \leftarrow$ Adjoint Equations [1 EFS] ★
  $r_j^0 = \frac{dF}{db_j} \leftarrow$ Gradient Expression
  $db_j^0 \leftarrow \text{init}(0)$
  $p_j \leftarrow -r_j^0$
  $m \leftarrow 0$
  while $r^m \neq 0$ **and** $m \leq M_{CG}$ **do**
    $\frac{dU_n}{db_j} p_j \leftarrow$ DD (Flow Equations) [1 EFS] ★
    $\frac{d\Psi_n}{db_j} p_j \leftarrow$ DD (Adjoint Equations) [1 EFS] ★
    $w_i = \frac{d^2F}{db_i db_j} p_j \leftarrow$ Hessian Expression
    $\eta \leftarrow \frac{r_i^m r_i^m}{p_j w_j}$
    $db_j^{m+1} \leftarrow db_j^m + \eta p_j$
    $r_j^{m+1} \leftarrow r_j^m + \eta w_j$
    $\beta \leftarrow \frac{r_i^{m+1} r_i^{m+1}}{r_j^m r_j^m}$
    $p_j \leftarrow -r_j^{m+1} + \beta p_j$
    $m \leftarrow m + 1$
  end while
  $b_j \leftarrow b_j + db_j$
  $k \leftarrow k + 1$
end while

Total Cost= 2+2M$_{CG}$ << N

★ ★
★ ★
★
★
⋮

Truncated Newton Methods (Εξήγηση)

$$\frac{\delta F}{\delta b_i}=\frac{\partial F}{\partial b_i}+\Psi_m\frac{\partial R_m}{\partial b_i}$$

$$\frac{\delta R_m}{\delta b_i}=\frac{\partial R_m}{\partial b_i}+\frac{\partial R_m}{\partial U_k}\frac{\delta U_k}{\delta b_i}=0.$$

$$R_k^{\Psi}=\frac{\partial F}{\partial U_k}+\Psi_m\frac{\partial R_m}{\partial U_k}=0$$

$$\frac{\delta F}{\delta b_i}=\frac{\partial F}{\partial b_i}+\Psi_m\frac{\partial R_m}{\partial b_i}$$

$$\frac{\delta^2 F}{\delta b_i\delta b_j}s_j=\frac{\partial^2 F}{\partial b_i\partial b_j}s_j+\frac{\partial^2 F}{\partial b_i\partial U_k}\underline{\left(\frac{\delta U_k}{\delta b_j}s_j\right)}+\Psi_m\frac{\partial^2 R_m}{\partial b_i\partial b_j}s_j+\Psi_m\frac{\partial^2 R_m}{\partial b_i\partial U_k}\underline{\left(\frac{\delta U_k}{\delta b_j}s_j\right)}$$
$$+\underline{\left(\frac{\delta \Psi_m}{\delta b_j}s_j\right)}\frac{\partial R_m}{\partial b_i}.$$

ΣΥΜΠΕΡΑΣΜΑ: Χρειάζομαι γινόμενα των κλίσεων των πρωτευουσών και των συζυγών μεταβλητών με κάποιο διάνυσμα και όχι τις κλίσεις αυτές καθαυτές!!!!

Truncated Newton Methods (Εξήγηση)

<u>Υπολογισμός του :</u> $\left(\frac{\delta U_k}{\delta b_j} s_j\right)$

$$\frac{\partial R_m}{\partial b_j} s_j + \frac{\partial R_m}{\partial U_k}\underline{\left(\frac{\delta U_k}{\delta b_j} s_j\right)} = 0,$$

… με κόστος περίπου αυτό του να λυθούν οι εξισώσεις ροής.

<u>Υπολογισμός του :</u>

$$\frac{\partial F}{\partial U_k} + \Psi_m \frac{\partial R_m}{\partial U_k} = 0$$

$$\frac{\partial^2 F}{\partial U_n \partial b_j} s_j + \frac{\partial^2 F}{\partial U_n \partial U_k}\left(\frac{\delta U_k}{\delta b_j} s_j\right) + \Psi_m \frac{\partial^2 R_m}{\partial U_n \partial b_j} s_j + \Psi_m \frac{\partial^2 R_m}{\partial U_n \partial U_k}\left(\frac{\delta U_k}{\delta b_j} s_j\right) + \underline{\left(\frac{\delta \Psi_m}{\delta b_j} s_j\right)} \frac{\partial R_m}{\partial U_n} = 0$$

… με κόστος περίπου αυτό του να λυθούν οι εξισώσεις ροής.

Πολύ Βασικό:

Ότι παρουσιάστηκε για τη Διακριτή Συζυγή Μέθοδο, εφαρμόζεται εξίσου και με τη Συνεχή Συζυγή Μέθοδο.

Απλώς, προτιμήθηκε η παρουσίαση να γίνει με τη Διακριτή Μέθοδο, γιατί είναι περισσότερο σύντομη και εποπτική!