Gradient Descent για Smooth Συναρτήσεις

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

Θεώρημα (Για σφάλμα ε θέλουμε $O\left(\frac{1}{\varepsilon}\right)$ βήματα)

Αν

- $f: \mathbb{R}^n \to \mathbb{R}$, κυρτή, β-smooth.
- $f(x_*) = min_{x \in K} f(x)$

τότε για να πάρουμε ένα $\hat{x} \in \mathbb{R}^n$, για το οποίο $f(\hat{x}) - f(x_*) \leq \varepsilon$,
αρκεί να τρέξουμε GD με $\eta = \frac{1}{\beta}$ για $O\left(\frac{\beta \|x_1 - x_*\|^2}{\varepsilon}\right)$ βήματα.

(x_1 είναι το σημείο εκκίνησης)

Απόδειξη

Την προηγούμενη φορά είδαμε οτι $f(x_{t+1}) \leq f(x_t) - \frac{1}{2\beta} \|\nabla f(x_t)\|^2$

$$f(x_*) \geq f(x_t) + \nabla f(x_t) \cdot (x_* - x_t)$$

$$\|\nabla f(x_t)\| \cdot \|x_t - x_*\| \geq -\nabla f(x_t) \cdot (x_* - x_t) \geq f(x_t) - f(x_*)$$

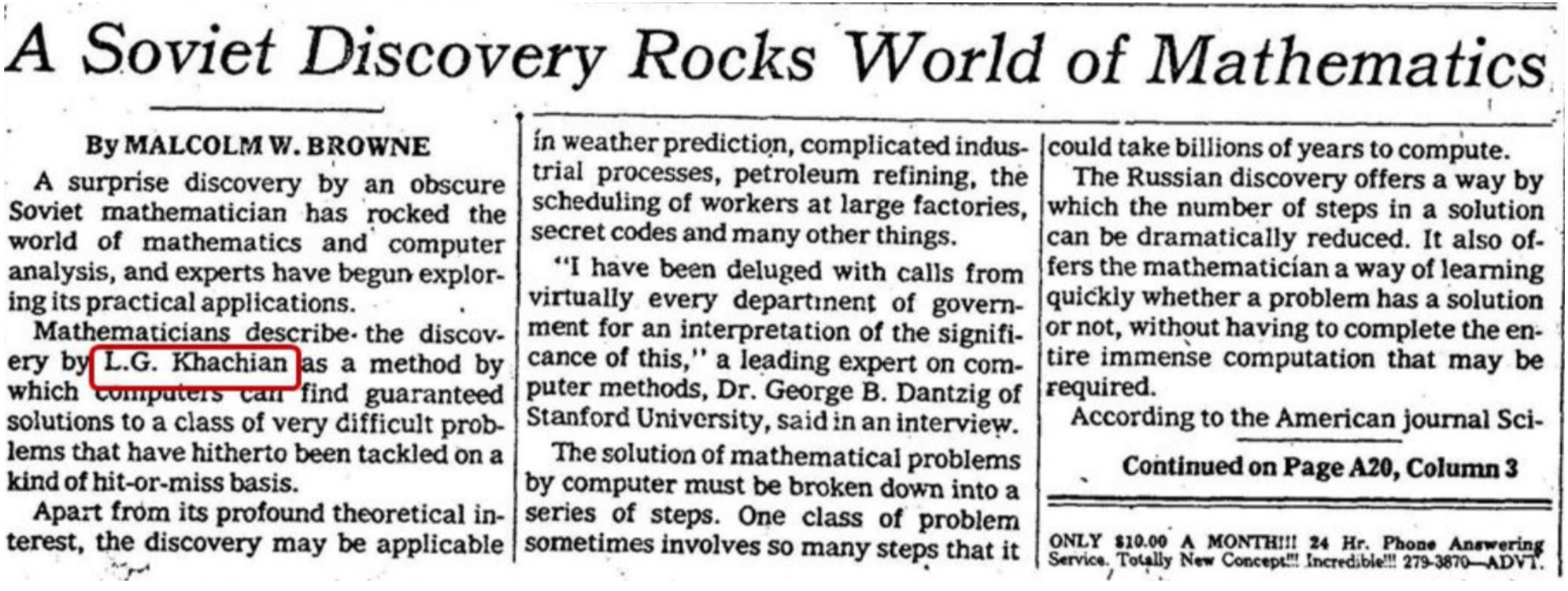
$$\|\nabla f(x_t)\| \geq \frac{f(x_t) - f(x_*)}{\|x_t - x_*\|} \geq \frac{f(x_t) - f(x_*)}{\|x_1 - x_*\|}$$

$$f(x_{t+1}) - f(x_*) \leq f(x_t) - f(x_*) - \frac{(f(x_t) - f(x_*))^2}{C}$$

$$\Delta_{t+1} \leq \Delta_t - \frac{\Delta_t{}^2}{C}$$

Γραμμικός Προγραμματισμός σε Πολυωνυμικό Χρόνο

Εξώφυλλο New York Times, 1979



A Soviet Discovery Rocks World of Mathematics

By MALCOLM W. BROWNE

A surprise discovery by an obscure Soviet mathematician has rocked the world of mathematics and computer analysis, and experts have begun exploring its practical applications.

Mathematicians describe the discovery by L.G. Khachian as a method by which computers can find guaranteed solutions to a class of very difficult problems that have hitherto been tackled on a kind of hit-or-miss basis.

Apart from its profound theoretical interest, the discovery may be applicable in weather prediction, complicated industrial processes, petroleum refining, the scheduling of workers at large factories, secret codes and many other things.

"I have been deluged with calls from virtually every department of government for an interpretation of the significance of this," a leading expert on computer methods, Dr. George B. Dantzig of Stanford University, said in an interview.

The solution of mathematical problems by computer must be broken down into a series of steps. One class of problem sometimes involves so many steps that it could take billions of years to compute.

The Russian discovery offers a way by which the number of steps in a solution can be dramatically reduced. It also offers the mathematician a way of learning quickly whether a problem has a solution or not, without having to complete the entire immense computation that may be required.

According to the American journal Sci-

Continued on Page A20, Column 3

ONLY $10.00 A MONTH!!! 24 Hr. Phone Answering Service. Totally New Concept!!! Incredible!!! 279-3870—ADVT.

Ελλειψοειδή: Ισοδύναμοι Ορισμοί

Ορισμός (υπενθύμιση)

Ένα $E \subseteq \mathbb{R}^n$ είναι ελλειψοειδές αν $E = \{x \in \mathbb{R}^n : (x - x_0)^T \mathrm{A} (x - x_0) \leq 1\}$

για κάποιο $\mathrm{A} \succ 0, x_0 \in \mathbb{R}^n$

Θεώρημα (ισοδύναμοι ορισμοί)

1. $E \subseteq \mathbb{R}^n$ είναι ελλειψοειδές αν $E = \{x \in \mathbb{R}^n : (x - x_0)^T \mathrm{A}^{-1}(x - x_0) \leq 1\}$

για κάποιο $\mathrm{A} \succ 0, x_0 \in \mathbb{R}^n$

2. Έστω $\mathrm{A} \succ 0, x_0 \in \mathbb{R}^n$. Τότε,

$$\{x \in \mathbb{R}^n : (x - x_0)^T \mathrm{A}^{-1}(x - x_0) \leq 1\} = \mathrm{A}^{1/2}\left(\overline{\mathrm{B}(0,1)}\right) + x_0$$

3. $E \subseteq \mathbb{R}^n$ είναι ελλειψοειδές $\leftrightarrow$ $E = M\left(\overline{\mathrm{B}(0,1)}\right) + x_0$

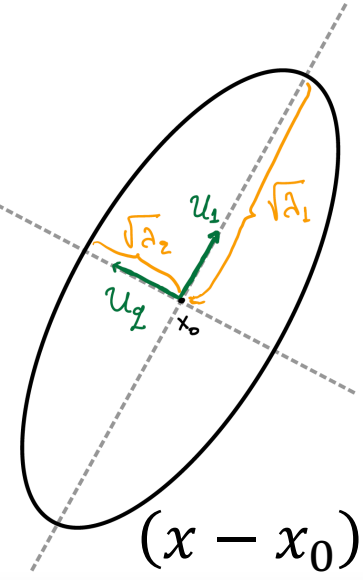
για κάποιο $\mathrm{M} \in \mathbb{R}^{n \times n}$ αντιστρέψιμο, και $x_0 \in \mathbb{R}^n$

Απόδειξη για το 1.

1. $E \subseteq \mathbb{R}^n$ είναι ελλειψοειδές αν $E = \{x \in \mathbb{R}^n : (x - x_0)^T A^{-1}(x - x_0) \leq 1\}$

για κάποιο $A \succ 0, x_0 \in \mathbb{R}^n$

Έστω $A \succ 0, x_0 \in \mathbb{R}^n$. Τότε, $A = U\Lambda U^T$, $\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$, $\lambda_i > 0$



$$A^{-1} = U\Lambda^{-1}U^T,\ \Lambda^{-1} = \begin{bmatrix} 1/\lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\lambda_n \end{bmatrix}$$

$$\begin{aligned}(x - x_0)^T A^{-1}(x - x_0) &= (x - x_0)^T U\Lambda^{-1}U^T(x - x_0) \\ &= (x - x_0)^T U(\Lambda^{-1})^{1/2}(\Lambda^{-1})^{1/2}U^T(x - x_0) \\ &= \left\|(\Lambda^{-1})^{1/2}U^T(x - x_0)\right\|^2 = \sum_{i=1}^n \left(\frac{u_i \cdot (x - x_0)}{\sqrt{\lambda_i}}\right)^2 \leq 1\end{aligned}$$

Απόδειξη για το 2.

Πρίν την απόδειξη:

Έστω U ορθοκανονικός. Τί κάνει η συνάρτηση $x \longmapsto Ux$ γεωμετρικά ?

Πρόταση: Έστω U ορθοκανονικός. Τότε η $x \longmapsto Ux$ διατηρεί 1) τις νόρμες ✓
2) τις αποστάσεις ✓

Απόδειξη: $\|Ux\|^2 = x^T U^T Ux = x^T x = \|x\|^2$

$$\|Ux - Uy\|^2 = \|U(x-y)\|^2 = \|x-y\|^2$$

Απόδειξη για το 2.

Έστω $A \succ 0, x_0 \in \mathbb{R}^n$. Θα δείξουμε ότι

$$\{x \in \mathbb{R}^n : (x - x_0)^T A^{-1}(x - x_0) \leq 1\} = A^{1/2}\overline{B(0,1)} + x_0$$

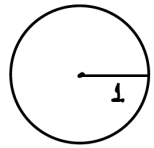
$$(A = U\Lambda U^T)$$

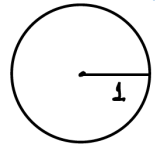
$$\leftrightarrow \left\{ \begin{array}{c} \text{Ελλειψοειδές} \\ \text{με άξονες } u_1, \dots, u_n \\ \text{μήκη αξόνων } \sqrt{\lambda_1}, \dots, \sqrt{\lambda_n} \\ \text{κέντρο το } 0 \end{array} \right\} + x_0 = U\Lambda^{1/2}U^T\overline{B(0,1)} + x_0$$

$$\leftrightarrow \left\{ \begin{array}{c} \text{Ελλειψοειδές} \\ \text{με άξονες } u_1, \dots, u_n \\ \text{μήκη αξόνων } \sqrt{\lambda_1}, \dots, \sqrt{\lambda_n} \\ \text{κέντρο το } 0 \end{array} \right\} = U\Lambda^{1/2}U^T\overline{B(0,1)}$$

Έστω $A \succ 0, x_0 \in \mathbb{R}^n$. Θα δείξουμε ότι

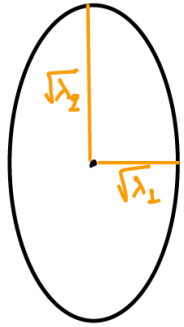
$$\left\{\begin{array}{c} \text{Ελλειψοειδές} \\ \text{με άξονες } u_1, \dots, u_n \\ \text{μήκη αξόνων } \sqrt{\lambda_1}, \dots, \sqrt{\lambda_n} \\ \text{κέντρο το } 0 \end{array}\right\} = U\Lambda^{1/2}U^T\left(\overline{\mathrm{B}(0,1)}\right)$$

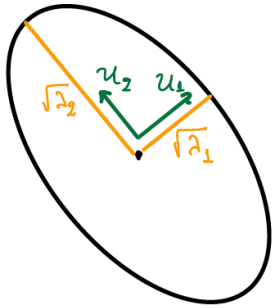




<u>1^{η} Περίπτωση:</u>

$$Rotation_1 \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix} Rotation_2 \left(\overline{\mathrm{B}(0,1)}\right)$$





Απόδειξη για το 3.

Θα δείξουμε ότι

$E \subseteq \mathbb{R}^n$ είναι ελλειψοειδές $\leftrightarrow$ $E = A\left(\overline{\mathrm{B}(0,1)}\right) + x_0$

για κάποιο $A \in \mathbb{R}^{n\times n}$ αντιστρέψιμο, και $x_0 \in \mathbb{R}^n$

Θεώρημα (Singular Value Decomposition - SVD)

Έστω $A \in \mathbb{R}^{n\times n}$. Υπάρχουν $U, V, \Sigma \in \mathbb{R}^{n\times n}$

- U, V ορθοκανονικοί
- $\Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n \end{bmatrix}, \quad \forall i,\ \sigma_i \geq 0$

και $A = \mathrm{U}\Sigma V^T$.

Αν A αντιστρέψιμος, τότε $\forall i,\ \sigma_i > 0$

Όγκος στον $\mathbb{R}^d$

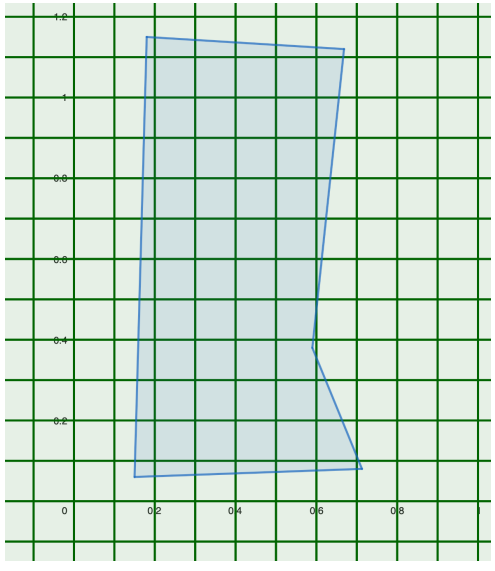
- Ονομάζουμε «κουτί» κάθε σύνολο $[a_1, b_1]\times[a_2, b_2]\times\cdots\times[a_d, b_d]$
- Ορίζουμε $vol(\,[a_1, b_1]\times[a_2, b_2]\times\cdots\times[a_d, b_d]\,) \coloneqq \prod_{i=1}^{d}(b_i - a_i)$
- Για $\mathrm{N} \in \mathbb{N}$, ορίζουμε το

$$\Pi\lambda\acute{\varepsilon}\gamma\mu\alpha_{\mathrm{N}} \coloneqq \left\{ \left[\frac{i_1}{N}, \frac{i_1+1}{N}\right]\times\left[\frac{i_2}{N}, \frac{i_2+1}{N}\right]\times\cdots\times\left[\frac{i_d}{N}, \frac{i_d+1}{N}\right] : i_1, i_2, \ldots, i_d \in \mathbb{Z} \right\}$$

- Έστω $S \subseteq \mathbb{R}^d$ φραγμένο. Ορίζουμε:

$$vol(S) \coloneqq \lim_{N\to\infty} \sum_{\substack{\kappa ουτί \in \Pi\lambda\acute{\varepsilon}\gamma\mu\alpha_N \\ \kappa ουτί \subseteq S}} vol(\kappa ουτί)$$

$$= \lim_{N\to\infty} \frac{1}{N^d}\ \#\kappa ουτιών \subseteq \mathrm{S}\ \tau ου\ \Pi\lambda\acute{\varepsilon}\gamma\mu\alpha_{\mathrm{N}}$$



Όγκος στον $\mathbb{R}^d$

Θεώρημα Έστω $S \subseteq \mathbb{R}^d$ φραγμένο.

$$\forall A \in \mathbb{R}^{d\times d}, \qquad vol\big(A(S)\big) = |\det(A)| \cdot vol(S)$$

Απόδειξη Θα χρησιμοποιήσουμε τις ιδιότητες:

1. $\forall x_0 \in \mathbb{R}^d, \quad vol(S + x_0) = vol(S)$ (μεταφορά κατά x_0)

2. $\forall\, \mathrm{U} \in \mathbb{R}^{d\times d}$ ορθοκανονικό, $vol(\mathrm{U}(S)) = vol(S)$

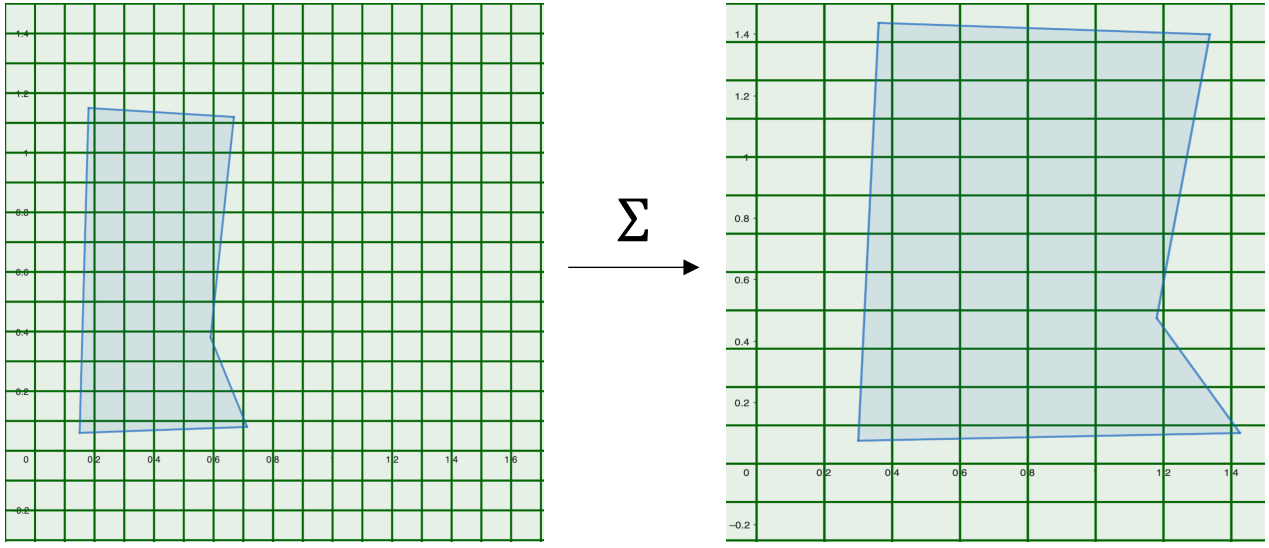
(περιστροφή ή περιστροφή ∘ ανάκλαση)

Όγκος στον $\mathbb{R}^d$

Απόδειξη (συνέχεια)

3. $\forall\, \Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_d \end{bmatrix}$ με $\sigma_i \geq 0$, έχουμε $vol(\Sigma(S)) = \left(\prod_{i=1}^{d} \sigma_i\right) vol(S)$

Παράδειγμα: $d = 2, \sigma_1 = 2, \sigma_2 = 1.25,\ S =$ το μπλέ σχήμα:



Ο Σ στο σχήμα εφαρμόζεται σε όλο το πλέγμα και στο S. Για κάθε κουτί η επιφάνειά του πολλαπλασιάστηκε με $\sigma_1\sigma_2$

Ιδέα για απόδειξη της 3 στον $\mathbb{R}^d$:

Για κάθε κουτί $= [a_1, b_1]\times[a_2, b_2]\times\cdots\ [a_d, b_d] \subseteq S$,

$\Sigma(\text{κουτί}) = [\sigma_1 a_1, \sigma_1 b_1]\times[\sigma_2 a_2, \sigma_2 b_2]\times\cdots\times[\sigma_d a_d, \sigma_d b_d]$

Όγκος στον $\mathbb{R}^d$

Θεώρημα Έστω $S \subseteq \mathbb{R}^d$ φραγμένο.

$$\forall A \in \mathbb{R}^{d\times d}, \qquad vol\big(A(S)\big) = |\det(A)| \cdot vol(S)$$

Απόδειξη (συνέχεια)

SVD: $A = \mathrm{U}\Sigma V^T$.

$$vol\big(A(S)\big) = vol\big(\mathrm{U}\Sigma V^T(S)\big) = \left(\prod_{i=1}^{d} \sigma_i\right) vol(S)$$

$$\det(A) = \det(U)\det(\Sigma)\det(V^T) = \pm\left(\prod_{i=1}^{d}\sigma_i\right)$$

(αν U ορθοκανονικός, τότε $\det(U) = \pm 1$. Απόδ: παίρνουμε det στο $UU^T = I$)

Όγκος στον $\mathbb{R}^d$

<u>Πόρισμα</u> Έστω $A \in \mathbb{R}^{d\times d}$ αντιστρέψιμος, $x_0 \in \mathbb{R}^d$. Το ελλειψοειδές $E = A\left(\overline{\mathrm{B}(0,1)}\right) + x_0$ έχει

$$vol(E) = |\det(A)| \cdot vol(\overline{\mathrm{B}(0,1)})$$