Ellipsoid methods with space scaling

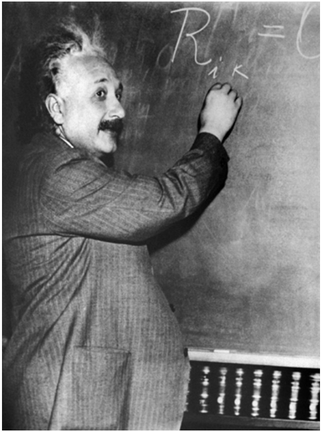
Petro Stetsyuk[1], Andreas Fischer[2], Olha Khomiak[1]

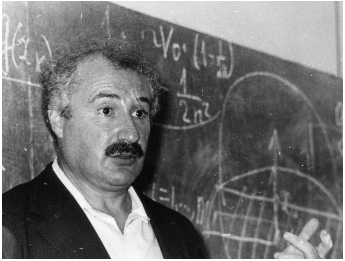
[1]V. M. Glushkov Institute of Cybernetics of NAS of Ukraine, Kyiv
[2]Institute of Numerical Mathematics, Technische Universität Dresden

The Seminar "Theory of Optimal Solutions"
Kyiv, November 15, 2022

Dedicated to the memory of academician
Naum Zuselevich Shor
(to the 85th anniversary of his birth)

Einstein writes a message to Shor





to denote space dilation operator as $R_\alpha(\xi)$

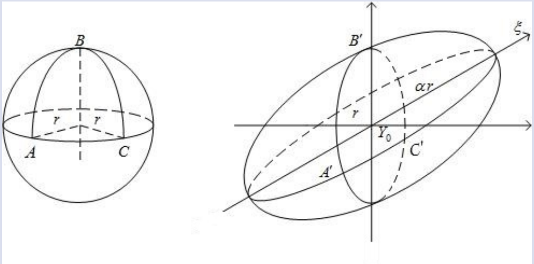
Shor's space dilation operator

Space dilation operator has the following form

$$R_\alpha(\xi) = I_n + (\alpha - 1)\xi\xi^T, \quad \text{where} \quad \alpha > 1.$$

Here: α is the coefficient of space dilation in the normed direction $\xi \in \mathbb{R}^n$, $\|\xi\| = 1$; I_n is the identity $n \times n$-matrix.

Example in $\mathbb{R}^3$: ball (left) dilate to ellipsoid (right)



Outline

1. The history of the ellipsoid method
2. The idea of ellipsoid method
3. Ellipsoid methods with space scaling
4. $2d$-ellipsoid and r-algoritms

Content

The ellipsoid method was proposed:

1976 by **Yudin and Nemirovski** as a method of successive cutting-plane [1];

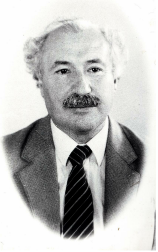
1977 by **Shor** as a variant of the method with space dilation in the direction of the subgradient [2].

1. Yudin D.B. and Nemirovski A.S. *Informational complexity and effective methods for the solution of convex extremal problems* // Ekonom. Mat. Metody, 12, No. 2 (1976).

2. Shor N.Z. *Cut-off method with space extension in convex programming problems* // Cybernetics, 13, No. 1 (1977).

Yudin and Shor „from the banks of the Dnipro“

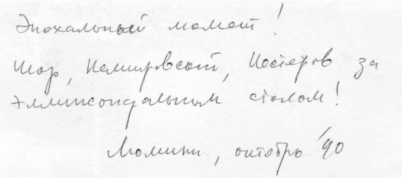
David Borisovich Yudin
born May 21, 1919
in Yekaterinoslav (today - Dnipro),
in 1941 graduated from
Dnepropetrovsk University



Naum Zuselevich Shor
born January 1, 1937
in Kyiv (city on the Dnipro),
in 1958 graduated from
Kyiv University

Epochal moment!

N. Shor,
A. Nemirovski,
Y. Nesterov at the ellipsoidal table!
October 1990


Эпохальный момент!
Шор, Немировский, Нестеров за эллипсоидальным столом!
Люмини, октябрь '90

XI ISMP, Bonn, August 23–27, 1982

Fulkerson Prizes for the ellipsoid method:

1. Grötchel M., Lóvasz L., Schrijver A., **1981**

2. Khachiyan L., **1979**, Yudin D., Nemirovski A., **1976**

Shor's plenary report:

„Generalized gradient methods of nondifferentiable optimization employing space dilatation operations“, published in [3].

3. Mathematical Programming: the state of art, Bonn, 1982 / *Bachem A., Grötchel M., Korte B. (eds.)* – Berlin: Springer-Verlag, 1983. – 655 p.

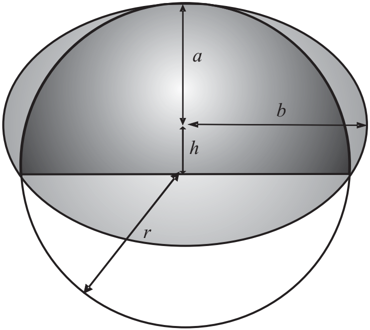
N. Shor (1982) and D. Yudin (1983)

N. Shor in Bonn (1982)

D. Yudin in Riga (1983)

Content

1d-ellipsoid and its properties



The 1d-ellipsoid $\mathcal{E}_n$, containing half of ball S_n in E^n, has parameters

$$b = \left(\alpha + \frac{1}{\alpha}\right)\frac{r}{2}, \quad h = \left(1 - \frac{1}{\alpha^2}\right)\frac{r}{2},$$

where $\alpha = \frac{b}{a}$ and r – radius of ball.

To transform $\mathcal{E}_n$ into a „new“ ball we have to dilate the space with coefficient $\alpha = \frac{b}{a}$, $\alpha > 1$.

The ratio of $\mathcal{E}_n$ to S_n volumes equals

$$q(n) = \frac{\text{vol}(\mathcal{E}_n)}{\text{vol}(S_n)} = \frac{1}{\alpha}\left(\frac{b}{r}\right)^n = \frac{1}{\alpha}\left(\frac{1}{2}\left(\alpha + \frac{1}{\alpha}\right)\right)^n.$$

Why ellipsoid method converges?

The ratio of $\mathcal{E}_n$ to S_n volumes equals

$$q(n) = \frac{\mathrm{vol}(\mathcal{E}_n)}{\mathrm{vol}(S_n)} = \frac{1}{\alpha}\left(\frac{1}{2}\left(\alpha + \frac{1}{\alpha}\right)\right)^n .$$

If coefficient α is such that $\alpha + 1/\alpha < 2\sqrt[n]{\alpha}$, then ratio $q(n) < 1$ thus volume of ellipsoid localizing searched point x^* shrinks with rate of geometric progression with ratio $q(n)$.

In Yudin-Nemirovski-Shor ellipsoid method

$q(n) \leq 1 - \frac{1}{2n}$ and is implemented with $\alpha = \sqrt{\frac{n+1}{n-1}}$.

Problem and stop criterion

Problem to solve:

for convex function $f(x)$, $x \in \mathbb{R}^n$ find point x^*_ε,
subject to $f(x^*_\varepsilon) - f^* \leq \varepsilon$, where $f^* = f(x^*)$.

Method input parameter:

$\varepsilon > 0$ – desired accuracy for finding $f_\varepsilon = f(x_\varepsilon)$.

Notation:

$g(x_k)$ – subgradient of $f(x)$ at x_k.

The B-form of algorithm **emshor**(x_0, r_0, ε)

Step 0. Choose $x_0 \in \mathbb{R}^n$, $r_0 > 0$, $\varepsilon > 0$, $\|x_0 - x^*\| \le r_0$.

Set $B_0 := I_n \in \mathbb{R}^{n \times n}$ (denoting the identity matrix) and $k := 0$.

Step 1. If $\left\|B_k^\top g(x_k)\right\| r_k \le \varepsilon$, then STOP: $k^* := k$, $x_\varepsilon^* := x_k$.

Step 2. Compute

$$x_{k+1} := x_k - \frac{r_k}{n+1} B_k \xi_k, \quad \text{where} \quad \xi_k := \frac{B_k^\top g(x_k)}{\left\|B_k^\top g(x_k)\right\|}.$$

Step 3. Update

$$B_{k+1} := B_k + \left(\sqrt{\frac{n-1}{n+1}} - 1\right) (B_k \xi_k)\, \xi_k^\top, \quad r_{k+1} := \frac{n}{\sqrt{n^2-1}} r_k.$$

Step 4. Set $k := k+1$ and go to Step 1.

Content

Algorithm **em22b**(λ, x_0, r_0, ε)

Step 0. Choose $\lambda > 0$, $x_0 \in \mathbb{R}^n$, $r_0 > 0$, $\varepsilon > 0$, $\|x_0 - x^*\| \leq r_0$.

Set $B_0 := I_n \in \mathbb{R}^{n \times n}$ (denoting the identity matrix) and $k := 0$.

Step 1. If $\left\|B_k^\top g(x_k)\right\| r_k \leq \varepsilon$, then STOP: $k^* := k$, $x_\varepsilon^* := x_k$.

Step 2. Compute

$$x_{k+1} := x_k - \frac{r_k}{n+1} B_k \xi_k, \quad \text{where} \quad \xi_k := \frac{B_k^\top g(x_k)}{\left\|B_k^\top g(x_k)\right\|}.$$

Step 3. Update

$$B_{k+1} := \lambda \left(B_k + \left(\sqrt{\frac{n-1}{n+1}} - 1 \right) (B_k \xi_k) \, \xi_k^\top \right), r_{k+1} := \frac{1}{\lambda} \frac{n}{\sqrt{n^2 - 1}} r_k.$$

Step 4. Set $k := k + 1$ and go to Step 1.

Theorem 1

For any $(\lambda, x_0, r_0, \varepsilon) \in (0, \infty) \times \mathbb{R}^n \times (0, \infty) \times (0, \infty)$, algorithm **em22b** is well-defined and generates a sequence $\{x_k\}_{k=0}^{k^*}$. With $A_k := B_k^{-1}$, it holds that

$$\|A_k(x_k - x^*)\| \leq r_k \quad \text{for} \quad k = 0, 1, 2, \ldots, k^*.$$

Theorem 2

There is $k^* \in \mathbb{N}$ so that algorithm **em22b** stops at Step 1 for $k = k^*$. For each k with $1 \leq k \leq k^*$, the ratio of the volumes of the ellipsoids E_k and E_{k-1} is a constant q_n with

$$q_n = \frac{vol(E_k)}{vol(E_{k-1})} = \sqrt{\frac{n-1}{n+1}} \left(\frac{n}{\sqrt{n^2-1}} \right)^n < \exp\left\{ -\frac{1}{2n} \right\} < 1.$$

Moreover, $f(x_{k^*}) - f^* \leq \varepsilon$ is satisfied.

Convergence rate of **em22b** (Shor, 1977)

for function

$$f(x) = \sum_{i=1}^{10} 2^{i-1}|x_i - 1|,\ x_0 = (0, \ldots, 0)^\top,\ r_0 = 10.$$

Shor, 1977, $\lambda = 1$, $b = \frac{n}{\sqrt{n^2-1}} r$

ε	f_ε^*	k^*	$\|B_k\|^*$	r_k^*
1.0e-02	4.8e-05	2151	1.1e-08	4.9e+05
1.0e-04	2.2e-06	3124	6.4e-13	6.6e+07
1.0e-06	2.0e-09	4024	8.1e-17	6.1e+09
1.0e-07	6.9e-09	4474	9.0e-19	5.8e+10
1.0e-08	6.5e-10	**4827**	2.6e-20	3.4e+11

Convergence rate of **em22b** (Khachiyan, 1980)

Khachiyan, 1980, $\lambda = \frac{n}{\sqrt{n^2-1}}$, $b = r$

ε	f^*_ε	k^*	$\|B^*_k\|$	r^*_k
1.0e-06	2.0e-09	4024	4.9e-08	10
1.0e-07	6.9e-09	4474	5.2e-09	10
1.0e-08	6.5e-10	**4934**	5.0e-10	10

4. KHACHIAN L.G. *Polynomial algorithms in linear programming*, USSR Computational Mathematics and Mathematical Physics, Vol. 20, No. 1, 1980, pp. 53–72.

Convergence rate of **em22b** (Nemirovski, Yudin)

Nemirovski, Yudin, 1979, $\lambda = \sqrt[n]{\frac{n+1}{n-1}}$, $b = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2n}} \frac{n}{\sqrt{n^2-1}} r$

ε	f_ε^*	k^*	$\|B_k\|$	r_k
1.0e-06	2.0e-09	4024	2.8e+01	1.8e-08
1.0e-07	6.9e-09	**4490**	2.8e+01	1.7e-09
1.0e-08	6.5e-10	**4953**	2.8e+01	1.7e-10

5. Nemirovski A., Yudin D. *Problem Complexity and Method Efficiency in Optimization*, John Wiley, New York, 1983, 388 p., translated from book published by Nauka, Moscow, 1979

Two remarks for the algorithm **em22b**

Though the three versions of **em22b** are equivalent,

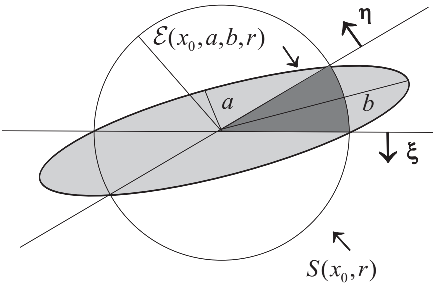
we observe slight differences in the number of iterations for $\varepsilon \in \{10^{-7}, 10^{-8}\}$ due to accumulation of numerical errors. A study of such effects for different f, n, and ε is intended.

Algorithm **em22b**

can be accelerated by tighter ellipsoidal approximations and applied to convex programs or saddle point problems for convex-concave functions.

Content

Minimum volume $2d$-ellipsoid



The transformation into the ball requires dilation

in the direction $\frac{\xi-\eta}{\|\xi-\eta\|}$

with $\alpha_1 = \frac{1}{\sqrt{1+(\xi,\eta)}} > 1$

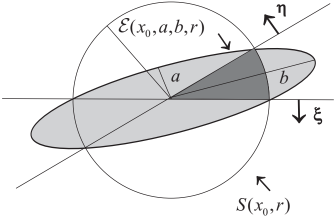
and compression

in the direction $\frac{\xi+\eta}{\|\xi+\eta\|}$

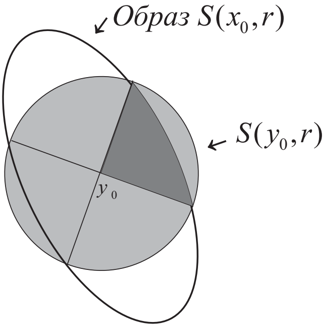
with $\alpha_2 = \frac{1}{\sqrt{1-(\xi,\eta)}} < 1.$

$$q = \frac{\text{vol}(\mathcal{E}(x_0,a,b,r))}{\text{vol}(S(x_0,r))} = \left(\frac{a}{r}\right)\left(\frac{b}{r}\right) = \sqrt{1-(\xi,\eta)^2}.$$

$2d$-ellipsoid before and after dilation



Minimum volume $2d$-ellipsoid



in two times transformed space turns into the ball with radius r

One-rank ellipsoidal operator

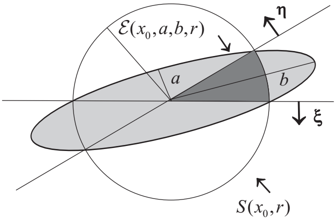
One-rank ellipsoidal operator is linear operator

$$T_1(\xi,\eta) = I - \frac{1}{1-(\xi,\eta)^2}\left(\left(1-\sqrt{1-(\xi,\eta)^2}\right)\eta - (\xi,\eta)\xi\right)\eta^T,$$

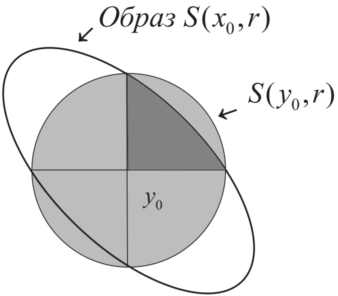
mapping from R^n to R^n. Here $\xi, \eta \in R^n$ – vectors, such that $\|\xi\| = 1, \|\eta\| = 1$ и $(\xi,\eta)^2 \neq 1$, I – identity $n \times n$-matrix.

6. Stetsyuk P.I. *Orthogonalizing linear operators in convex programming.* Cybern. and Systems Analysis. 54, 1997, pp. 386–401.

$2d$-ellipsoid before and after dilation



Minimum volume $2d$-ellipsoid



in transformed space
turns into the ball

Connection with r-algorithms

In transformed space the $2d$-ellipsoid turns into the ball, and images of vectors ξ and η are orthogonal.

It allows to "expand" a cone of suitable directions of the function decreasing for the subgradient process in a transformed space of variables, similar to how it is done in r-algorithms (Shor and Zhurbenko).

Space dilation is implemented in the direction of difference of two normed subgradients, and it is close to the direction of difference of subgradients if the subgradient norms are close.

Conclusion

Shor N.Z. and Gershovich V.I. in 1982 wrote

„Теория всего класса алгоритмов с растяжением пространства далека от совершенства. Нам кажется достаточно реалистичной целью – построение такого алгоритма, который по своей практической эффективности не уступал бы r-алгоритму и был столь же хорошо обоснован как метод эллипсоидов“.

We plan to achieve this goal

7. Стецюк П.И. *Методы эллипсоидов и r-алгоритмы*. Кишинэу, Эврика, 2014. 488 с.

References 2018 – 2022

- Stetsyuk P.I., Fesiuk O.V., Khomyak O.N. The generalized ellipsoid method. *Cybernetics and Systems Analysis*, V. 54, 2018, pp. 576–584.
- Izmailov A.F., Stetsyuk P.I., Fischer A. Emshor algoritm and its octave implementation, *Computer Mathematics*, No. 1, 2019, pp. 132–142.
- Stetsyuk P., Fischer A., Khomyak O. The generalized ellipsoid method and its implementation, *Optimization and Applications. OPTIMA 2019. Communications in Computer and Information Science*, Vol. 1145, 2020, pp. 355–370.
- Stetsyuk P.I., Fischer A., Khomiak O.M. Ellipsoid methods with space scaling. *Proceedings of the 8th International Conference on Control and Optimization with Industrial Applications*. 24–26 August, 2022, Baku, Azerbaijan. pp. 402–404.

Questions?

THANK YOU FOR YOUR ATTENTION!

e-mail: stetsyukp@gmail.com