

Ellipsoid methods with space scaling

Petro Stetsyuk¹, Andreas Fischer², Olha Khomiak¹

¹V.M. Glushkov Institute of Cybernetics,
Kyiv, Ukraine

²Institute of Numerical Mathematics,
Technical University of Dresden

The 8th International Conference on
Control and Optimization with Industrial Applications (COIA-2022)
Baku, August 24-26, 2022

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

ПОЕХАЛ В ГРЕЦИЮ, а чемодан остался в Венгрии...



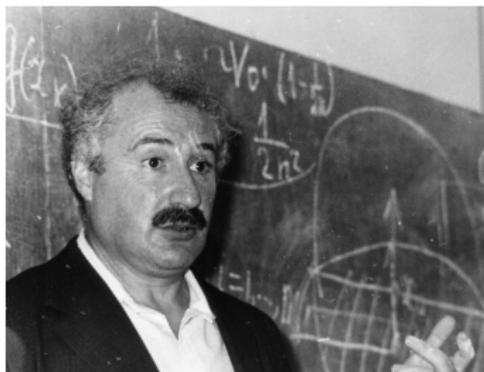
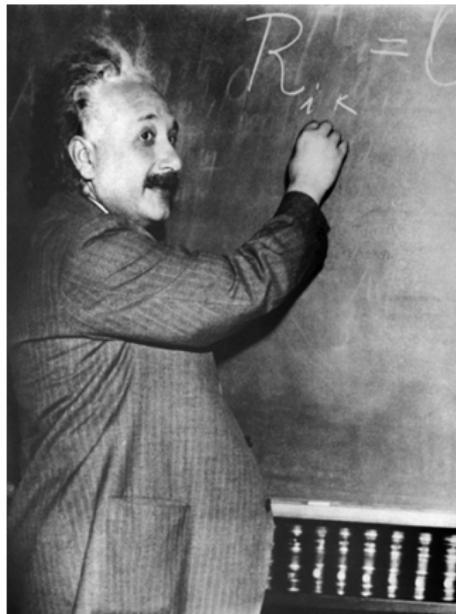
Академик НАН Украины профессор Наум Шор, родившийся 1 января 1937 года, имеет железную логику и ясный ум. Он 44 года без перерыва работает в Институте кибернетики им. В. М. Глушкова. Сегодня занимает там должность заведующего отделом методов решения сложных задач оптимизации.

Коллеги из Швейцарии пригласили ученого в музей Эйнштейна в Берне, чтобы сделать это фото и подчеркнуть их сходство

|| то обо мне писать? — засмутился Наум Зуселевич. — Я живу: с ра-

комился 40 лет назад, 1 января 1963 года. Звезды расположились так, что и она Ко-

Einstein writes a message to Shor



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

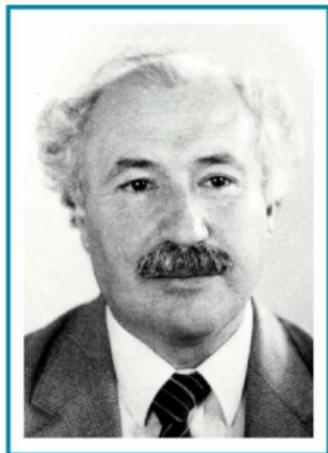
Outline

- 1 Shor's non-smooth optimization methods
- 2 Subgradient methods with space transformation
- 3 Others about N.Z. Shor
- 4 The history of the ellipsoid method
- 5 The idea of ellipsoid method
- 6 Ellipsoid methods with space scaling

Content

- 1 Shor's non-smooth optimization methods
- 2 Subgradient methods with space transformation
- 3 Others about N.Z. Shor
- 4 The history of the ellipsoid method
- 5 The idea of ellipsoid method
- 6 Ellipsoid methods with space scaling

Naum Z. Shor (1937–2006)



Shor founded scientific school of non-smooth optimization methods (Institute of Cybernetics, Kyiv, Ukraine)

In **1962** he developed the first subgradient method.

In **1969** he used the space dilation operator for acceleration of gradient methods convergences.

N.Z. Shor's methods

are still of a great theoretical and applied importance, and "**a key**" for solution of large-scale problems.

Shor's basic monographies

1. ШОР Н.З. *Методы минимизации недифференцируемых функций и их приложения*. Киев: Наукова думка, 1979.

English translation: SHOR N.Z. *Minimization Methods for Non-Differentiable Functions*. Berlin: Springer-Verlag, 1985.

2. SHOR N.Z. *Nondifferentiable optimization and polynomial problems*. Boston; Dordrecht; London: Kluwer Academic Publishers, 1998.

3. ШОР Н.З., СТЕЦЕНКО С.И. *Квадратичные экстремальные задачи и недифференцируемая оптимизация*. Киев: Наукова думка, 1989.

Shor's Three key Ideas

Sergienko I.V., Stetsyuk P.I.

On N.Z. Shor's three scientific ideas. *Cybernetics and Systems Analysis* 48, 2–16 (2012).

The paper is devoted to the 75th anniversary of N.Z. Shor.

This paper described Shor's three key ideas:

generalized gradient descent (1962),

the use of linear nonorthogonal space transformations to improve the conditionality of ravine functions (1969),

dual approach for finding bounds of the objective function in nonconvex quadratic models (1985).

Examples of the application of these ideas in methods and algorithms developed at the V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine are given.

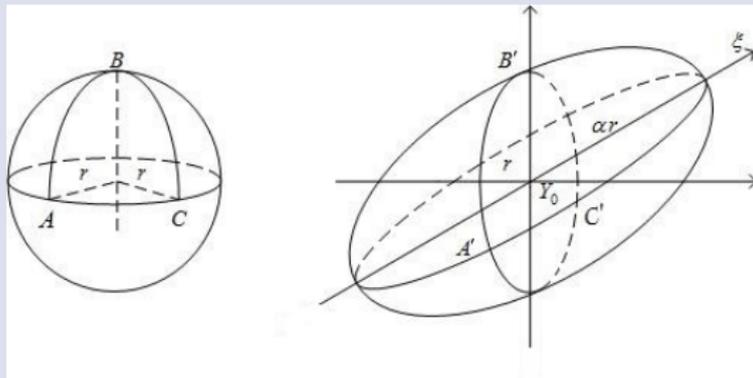
Shor's space dilation operator

Space dilation operator has the following form

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

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)



Content

- 1 Shor's non-smooth optimization methods
- 2 Subgradient methods with space transformation**
- 3 Others about N.Z. Shor
- 4 The history of the ellipsoid method
- 5 The idea of ellipsoid method
- 6 Ellipsoid methods with space scaling

Subgradient methods with space transformation

purposed to solve the following problem

$$f^* = f(x^*) = \min_{x \in \mathbb{R}^n} f(x),$$

where $f(x)$ – a convex function (smooth, non-smooth)

We have: x_0 – starting point, B_0 – $n \times n$ -matrix,

Iterations $k=1, 2, \dots$ have the following form

$$x_{k+1} = x_k - h_k B_k \frac{B_k^T g_f(x_k)}{\|B_k^T g_f(x_k)\|}, \quad B_{k+1} = B_k T_k, \quad (\text{Shor69})$$

where h_k – step-size, $g_f(x_k)$ – a subgradient of function $f(x)$ at the point x_k , T_k – $n \times n$ -matrix.

The most known methods

1. **r -algorithms** (robust to accumulation of errors);
2. **ellipsoid methods** (convergence - geometric progression);
3. **subgradient methods with space transformation**
(use Fejer-type steps, Polyak's steps and others).

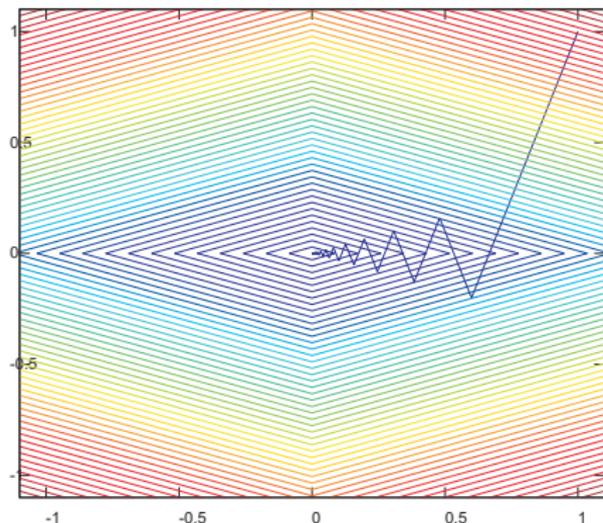
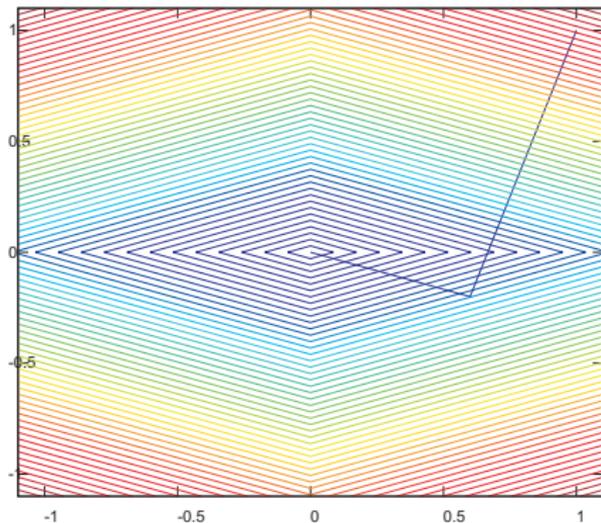
The methods have accelerated convergence

for convex ravine functions (smooth and non-smooth).

Implementations of the methods in Octave:

1. **ralgb5a** (30 rows),
2. **emshor** (15),
3. **amsg2p** (25).

Accelerated convergence: piecewise linear function

1. Trajectory for **ams** method2. ...accelerated **ams_{g2}** method

for ravine function $f_1(x_1, x_2) = |x_1| + 10|x_2|$, $x_0 = (1, 1)$.

Implementations of r -algorithms

were used for solving:

1. large-scale block optimization problems with various decomposition schemes;
2. minimax and matrix optimization problems;
3. for calculating dual bounds in multiextremal and combinatorial optimization problems;
4. ... and others

They were core for application packages, C and SA

PLANER (1983, **19**, 362–382), DISNEL (1991, **27**, 354–366).

Content

- 1 Shor's non-smooth optimization methods
- 2 Subgradient methods with space transformation
- 3 Others about N.Z. Shor**
- 4 The history of the ellipsoid method
- 5 The idea of ellipsoid method
- 6 Ellipsoid methods with space scaling

Boris Polyak „Introduction to Optimization“ (1987)



Boris Teodorovich says:

5.3 The Subgradient Method, p. 138

„The fundamental algorithms for minimizing smooth functions, the gradient as well as Newton’s algorithms, are based on linear or quadratic approximation of the function given by the first terms of a Taylor series. However this method is unfeasible for nondifferentiable functions, for such a function cannot be well approximated either by a linear or by a quadratic function.“

Boris Polyak „Introduction to Optimization“ (1987)

5.3 The Subgradient Method, p. 139

„Methods for minimizing nonsmooth functions cannot be further developed without new, innovative techniques. N.Z. Shor suggests – however surprisingly – a direct analog of the gradient method, with the gradient replaced by an arbitrary subgradient of the function $f(x)$:

$$x_{k+1} = x_k - \gamma_k g_f(x_k). \quad (3)$$

... the values of the function in method (3) cannot decrease monotonically. In this case, however, another function, viz. the distance to the minimum point, decreases monotonically. This is the key idea of the subgradient method (3).“

Stephen Boyd, Stanford University



1. BOYD, S., BARRATT, C. *Linear Controller Design: Limits of Performance* (1991)

2. BOYD, S. AND ALL. *Linear Matrix Inequalities in System and Control Theory* (1994)

3. BOYD, S., VANDENBERGHE, L. *Convex Optimization* (2004)

Boyd's letter to Shor (April 15, 2005)

Dear Professor Shor,

We have never met, but your work has very much influenced me for many years now. I started with your small 1985 Springer book on subgradient methods, which I read as a PhD student. I recently read your newer book on nondifferentiable optimization (1998), which I enjoyed very much.

*I'm enclosing copies of the three books I've written. **The first** concerns the design of linear controllers via convex optimization; **the second** is on linear matrix inequalities; and **the third** one is a basic textbook on convex optimization. [...] I hope you can see your strong influence in all of these books.*

*With the best regards,
Stephen P. Boyd*

„Optimization Methods and Software“ (2008), dedicated to the memory of N. Shor



B. Mordukhovich



M. Solodov



M. Todd

"In 1972, Shor introduced the fundamental generalized differential notion for locally Lipschitzian functions, which he called "the set of almost-gradients"..."

B. Mordukhovich, M. Solodov, M. Todd (2008)

„...It was defined as the collection of limiting points of the usual gradients of the Lipschitz continuous function in question, which is differentiable almost everywhere by the classical Rademacher theorem. This limiting set was later widely used, under the name of B-gradient and B-Jacobian for the case of vector functions, in developing nonsmooth versions of Newton's method. It is worth mentioning that in the same paper of 1972, Shor also introduced and utilized the convex hull of the set of almost-gradients, which he called the 'set of generalized almost gradients'. This latter set was subsequently rediscovered by Clarke and was widely used in nonsmooth optimization under the name of (Clarke's) generalized gradient for Lipschitzian functions.“

Content

- 1 Shor's non-smooth optimization methods
- 2 Subgradient methods with space transformation
- 3 Others about N.Z. Shor
- 4 The history of the ellipsoid method**
- 5 The idea of ellipsoid method
- 6 Ellipsoid methods with space scaling

The ellipsoid method was proposed:

- 1976 by **Yudin and Nemirovskii** 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 NEMIROVSKII 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
эллипсоидальной столом!
Москва, 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)

N. Shor in Azerbaijan (1982)

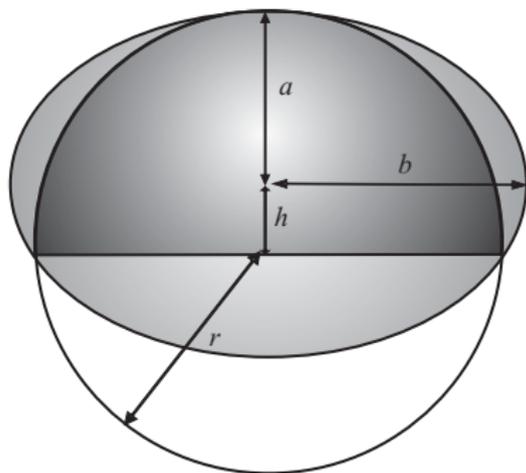


F. Akhmedov, K. Aida-zade, N. Shor, E. Nasibov (Baku, 1982)

Content

- 1 Shor's non-smooth optimization methods
- 2 Subgradient methods with space transformation
- 3 Others about N.Z. Shor
- 4 The history of the ellipsoid method
- 5 The idea of ellipsoid method**
- 6 Ellipsoid methods with space scaling

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 ellipsoids method converges?

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

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

If coefficient α is such that $\alpha + 1/\alpha < 2\sqrt{\alpha}$, 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

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

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

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

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

Step 2. Compute

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

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

- 1 Shor's non-smooth optimization methods
- 2 Subgradient methods with space transformation
- 3 Others about N.Z. Shor
- 4 The history of the ellipsoid method
- 5 The idea of ellipsoid method
- 6 Ellipsoid methods with space scaling**

Algorithm em22b($\lambda, x_0, r_0, \varepsilon$)

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

Step 2. Compute

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

Step 3. Update

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

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

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

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

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

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

Convergence rate of $\text{em22b}(\lambda, x_0, r_0, \varepsilon)$

for function

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

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

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

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

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

Conclusion

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.

References (1) – (3)

-  Izmailov A.F., Stetsyuk P.I., Fischer A., Emschor algorithm and its octave implementation, *Computer Mathematics*, No.1, 2019, pp.132-142.
-  Khachian L.G., Polynomial algorithms in linear programming, *USSR Computational Mathematics and Mathematical Physics*, Vol.20, No.1, 1980, pp.53-72.
-  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

References (4) – (6)

-  Sergienko I.V., Stetsyuk P.I., On N.Z. Shor's three scientific ideas. *Cybernetics and Systems Analysis*, Vol.48, No.1, 2012, pp.2–16.
-  Shor N.Z., Utilization of the operation of space dilation in the minimization of convex functions, *Cybernetics and Systems Analysis*, Vol.6, No.1, 1972, pp.7-15, translated from *Kibernetika*, Vol.6, No.1, 1970, pp.6-12.
-  Shor N.Z., Cut-off method with space extension in convex programming problems, *Cybernetics*, Vol.13, No.1, 1977, pp.94-96.

References (7), (8)

-  Stetsyuk P., Fischer A., Khomyak O., The generalized ellipsoid method and its implementation, *Optimization and Applications. OPTIMA 2019. Communications in Computer and Information Science*, (M. Jaćimović, M. Khachay, V. Malkova, M. Posypkin, editors), Vol.1145, 2020, pp.355-370.
-  Yudin D.B., Nemirovski A.S., Informational complexity and effective methods of solution of convex extremal problems (in Russian), *Èkonomika i Matematicheskie Metody*, Vol.12, No.2, 1976, pp. 357-369.

Questions?

THANK YOU FOR YOUR ATTENTION!

e-mail: stetsyukp@gmail.com