

Article

Approximation of the Constant in a Markov-Type Inequality on a Simplex Using Meta-Heuristics

Grzegorz Sroka ¹  and Mariusz Oszust ^{2,*} 

¹ Department of Topology and Algebra, Rzeszow University of Technology, Powstancow Warszawy 12, 35-959 Rzeszow, Poland; gsroka@prz.edu.pl

² Department of Computer and Control Engineering, Rzeszow University of Technology, Wincentego Pola 2, 35-959 Rzeszow, Poland

* Correspondence: marosz@kia.prz.edu.pl

Abstract: Markov-type inequalities are often used in numerical solutions of differential equations, and their constants improve error bounds. In this paper, the upper approximation of the constant in a Markov-type inequality on a simplex is considered. To determine the constant, the minimal polynomial and pluripotential theories were employed. They include a complex equilibrium measure that solves the extreme problem by minimizing the energy integral. Consequently, examples of polynomials of the second degree are introduced. Then, a challenging bilevel optimization problem that uses the polynomials for the approximation was formulated. Finally, three popular meta-heuristics were applied to the problem, and their results were investigated.

Keywords: minimal polynomials; complex equilibrium measure; Markov's inequality; bilevel optimization; optimization methods



Citation: Sroka, G.; Oszust, M. Approximation of the Constant in a Markov-Type Inequality on a Simplex Using Meta-Heuristics. *Mathematics* **2021**, *9*, 264. <https://doi.org/10.3390/math9030264>

Academic Editor: Yeol Je Cho
Received: 20 December 2020
Accepted: 26 January 2021
Published: 29 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Markov-type polynomial inequalities (or inverse inequalities) [1–6], similarly to Bernstein-type inequalities [1,5–9], are often found in many areas of applied mathematics, including popular numerical solutions of differential equations. Proper estimates of optimal constants in both types of inequalities can help to improve the bounds of numerical errors. They are often used in the error analysis of variational techniques, including the finite element method or the discontinuous Galerkin method used for solving partial differential equations (PDEs) [10]. Moreover, Markov's inequality can be found in constructions of polynomial meshes, as it is a representative sampling method, and is used as a tool for studying the uniform convergence of discrete least-squares polynomial approximations or for spectral methods for the solutions of PDEs [11–13]. Generally, finding exact values of optimal constants in a given compact set E in \mathbb{R}^N is considered particularly challenging. In this paper, an approach to determining the upper estimation of the constant in a Markov-type inequality on a simplex is proposed. The method aims to provide a greater value than that estimated by L. Bialas-Ciez and P. Goetgheluck [4].

For an arbitrary algebraic polynomial P of N variables and degree not greater than n , Markov's inequality is expressed as

$$\|gradP\|_E \leq Cn^2\|P\|_E, \quad (1)$$

where $gradP := \left(\frac{\partial P}{\partial x_1}, \dots, \frac{\partial P}{\partial x_N}\right)$, $|gradP| := \left(\sum_{j=1}^N \left(\frac{\partial P}{\partial x_j}\right)^2\right)^{1/2}$, $\|P\|_E = \sup\{|P(x)| : x \in E\}$, $C > 0$, E is a convex compact subset of \mathbb{K}^N (with non-void interior), and $\mathbb{K} \in \mathbb{R}$. The exponent 2 in (1) is optimal for a convex compact set E (with non-void interior) [2,4].

In 1974, D. Wilhelmsen [14] got an upper estimate for the constant $C < \frac{4}{\omega_E}$ in Markov's inequality (1), where ω_E is the width of E . The width of $E \in \mathbb{R}^N$ is defined as $\omega_E :=$

$\inf\{\text{dist}(Pl_1, Pl_2) \mid (Pl_1, Pl_2) \in G\}$ [4], where G is the set of all possible couples of parallel planes Pl_1 and Pl_2 such that E lies between Pl_1 and Pl_2 . Furthermore, for any convex compact set $E \subset \mathbb{R}^N$, the author proposed the hypothesis that $C = \frac{2}{\omega_E}$. M. Baran [3,7] and Y. Sarantopoulos [15] independently proved that Wilhelmsen's conjecture is true for centrally symmetric convex compact sets. In turn, L. Bialas-Cież and P. Goetgheluck [4] showed that the hypothesis is false in the simplex case and proved that the constant $C > 2.842534$.

The best constant in Markov's inequality (1) for a convex set was also studied by D. Nadzhmaddinow and Y. N. Subbotin [6] and by A. Kroó and S. Révész [16]. It is worth mentioning that for any triangles in \mathbb{R}^2 , A. Kroó and S. Révész [16] generalized the result for a simplex given in [4], considering an admissible upper bound for $C \leq \frac{\sqrt{10}}{\omega_E}$, and received an upper estimate value of constant C in (1): $C \leq \frac{s(\Delta)}{w(E)}$, where $s(\Delta) := \frac{2}{a} \sqrt{a^2 + b^2 + 2ab \cos \gamma'}$, with Δ being a triangle with the sides a, b, c and the angles α', β', γ' , respectively, such that $c \leq b \leq a$ and $\gamma' \leq \beta' \leq \alpha'$.

In the recent literature, many works have considered Markov inequalities. For example, S. Ozisik et al. [10] determined the constants in multivariate Markov inequalities on an interval, a triangle, and a tetrahedron with the L_2 -norm. They derived explicit expressions for the constants on all above-mentioned simplexes using orthonormal polynomials. It is worth noticing that exact values of the constants are the key for the correct derivation of both a priori and a posteriori error estimates in adaptive computations. F. Piazzon and M. Vianello [11] used the approximation theory notions of a polynomial mesh and the Dubiner distance in a compact set to determine error estimates for the total degree of polynomial optimization on Chebyshev grids of the hypercube. Additionally, F. Piazzon and M. Vianello [12] constructed norming meshes for polynomial optimization by using a classical Markov inequality on the general convex body in \mathbb{R}^N . A. Sommariva and M. Vianello [13] used discrete trigonometric norming inequalities on subintervals of the period. They constructed norming meshes with optimal cardinality growth for algebraic polynomials on the sections of a sphere, a ball, and a torus. Moreover, A. Kroó [17] studied the so-called optimal polynomial meshes and O. Davydov [18] presented error bounds for the approximation in the space of multivariate piecewise polynomials admitting stable local bases. In particular, the bounds apply to all spaces of smooth finite elements, and they are used in the finite element method for general fully nonlinear elliptic differential equations of the second order. L. Angermann and Ch. Heneke [19] proved error estimates for the Lagrange interpolation and for discrete L_2 -projection, which were optimal on the elements and almost optimal on the element edges. In another study, M. Bun and J. Thaler [5] reproved several Markov-type inequalities from approximation theory by constructing explicit dual solutions to natural linear programs. Overall, these inequalities are a basis for proofs of many approximated solutions that can be found in the literature.

In this paper, an approach to determining an approximation of the constant in Markov's inequality using introduced minimal polynomials is proposed. Then, the polynomials are used in the formulated bilevel optimization problem. Since many optimization techniques could be employed to find the constant, three popular meta-heuristics are applied and their results are investigated.

The rest of this paper is structured as follows. In Section 2, minimal polynomials are introduced. Then, in Section 3, their application in the approximation of the constant in a Markov-type inequality is considered. In Section 4, in turn, the approximation is defined as an optimization problem and the problem is solved using three meta-heuristics. Finally, in Section 5, the study is summarized and directions for future work are indicated.

2. Minimal Polynomials

The monic polynomial is defined as: $P(x) := x^\alpha + \sum_{\beta \prec \alpha} a_\beta x^\beta$, for each $x \in \mathbb{C}^N$, and with $a_\beta \in \mathbb{C}$ [20–22] (the coefficient of its leading monomial is equal 1), where $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{N}^N$, $|\alpha| := \alpha_1 + \dots + \alpha_N$ is its length and $x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \dots x_N^{\alpha_N}$; \prec indicates the graded lexicographic order, defined as: If $|\alpha| < |\alpha'|$, then $\alpha \prec \alpha'$ ($x^\alpha \prec x^{\alpha'}$), while if

$|\alpha| = |\alpha'|$ when $\alpha \prec \alpha'$ ($x^\alpha \prec x^{\alpha'}$), the first (starting from the left) non-zero coefficient of $\alpha' - \alpha$ is positive [21,22]. Polynomial $P(x)$ of the smallest deviation from zero on E with respect to the supremum norm is called the minimal polynomial (Chebyshev polynomial) for E [22–25].

Let E be a compact set in a complex plane that contains at least $|\alpha| + 1$ points. Based on the classical theorem of Tonelli [26], among the monic polynomials of degree $|\alpha|$, there exists only one polynomial $T_{|\alpha|}$ that minimizes the supremum norm on E . The $T_{|\alpha|}$ is called the Chebyshev polynomial of order $|\alpha|$. It is defined as $\|T_{|\alpha|}\|_E = \inf \left\{ \|x^{|\alpha|} + \sum_{i=0}^{|\alpha|-1} a_i x^i\|_E \right\}$, where the infimum is taken with respect to $(a_0, a_1, \dots, a_{|\alpha|-1}) \in \mathbb{C}^{|\alpha|}$. In other words, $T_{|\alpha|} - x^{|\alpha|}$ is the polynomial of best approximation to $x^{|\alpha|}$ on E . This polynomial is called the minimal (monic) polynomial of degree $|\alpha|$.

The complex equilibrium measure for a standard simplex in \mathbb{R}^N is calculated for $E = \{x \in \mathbb{R}^N : x_1, \dots, x_N \geq 0, x_1 + \dots + x_N \leq 1\} = S_N$, which is a simplex in \mathbb{R}^N [8]. Then, the plurisubharmonic extremal function on E , $u_E(z) = \log h(|z_1| + \dots + |z_N| + |z_1 + \dots + z_N - 1|)$, is used to calculate the (complex) equilibrium measure $\lambda_{S_N} = N! \text{vol}(B_N)(x_1 \cdot \dots \cdot x_N)^{-1/2} (1 - x_1 - \dots - x_N)^{-1/2} dx$, where $\text{vol}(B_N)$ is the usual volume of the closed-unit Euclidean ball in \mathbb{R}^N .

3. Estimation of the Constant C in a Markov-Type Inequality

Let S_2 be a simplex in \mathbb{R}^2 . Consequently,

$$S_2 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1, x_2 \geq 0, x_1 + x_2 \leq 1\}.$$

The normalized equilibrium measure S_2 , as a subset of \mathbb{C}^2 , is a two-dimensional Lebesgue measure with the weight $A(x_1, x_2) := \frac{1}{\sqrt{x_1} \sqrt{x_2} \sqrt{1-x_1-x_2}}$. For continuous functions f, g on S_2 , the scalar product is defined as

$$(f, g) = \iint_{S_2} f(x_1, x_2) g(x_1, x_2) A(x_1, x_2) dx_1 dx_2.$$

Then, a polynomial $Q(x_1, x_2) = x_1 + x_2 - 1/2$ is defined. It is the minimal polynomial for the simplex, and $P(x_1, x_2) = a_0 x_1^2 + a_1 x_2^2 + a_2 x_1 x_2 + a_3 x_1 + a_4 x_2 + a_5$, for some real numbers a_0, a_1, a_2, a_3, a_4 , and a_5 . Thus,

$$\begin{aligned} (P, Q) = \iint_{S_2} P(x_1, x_2) Q(x_1, x_2) A(x_1, x_2) dx_1 dx_2 &= a_0 x_{3,0} + a_1 x_{1,2} + a_2 x_{2,1} + a_3 x_{2,0} + a_4 x_{1,1} \\ &+ a_5 x_{1,0} + a_0 x_{2,1} + a_1 x_{0,3} + a_2 x_{1,2} + a_3 x_{1,1} + a_4 x_{0,2} + a_5 x_{0,1} - a_0 x_{2,0}/2 + a_1 x_{0,2}/2 \\ &- a_2 x_{1,1}/2 - a_3 x_{1,0}/2 - a_4 x_{0,1}/2 - a_5 x_{0,0}/2, \end{aligned}$$

where $x_{i,j}$ for $i, j \in \{0, 1, 2, 3\}$ are symbols given in Table 1.

Table 1. Symbols and their values.

SYMBOLS	$x_{0,0}$	$x_{1,0}$	$x_{0,1}$	$x_{1,1}$	$x_{1,2}$	$x_{2,1}$	$x_{2,0}$	$x_{0,2}$	$x_{3,0}$	$x_{0,3}$
VALUES	2π	$2\pi/3$	$2\pi/3$	$2\pi/15$	$2\pi/35$	$2\pi/35$	$2\pi/5$	$2\pi/5$	$2\pi/7$	$2\pi/7$

We want the polynomial P to be orthogonal to Q so that $(P, Q) = 0$. Substituting the values given in Table 1 into the expression above, $a_0/7 + a_1/7 + a_2/21 + a_3/5 + a_4/5 + a_5/3 = 0$ is obtained. Finally, $a_5 = -3a_0/7 - 3a_1/7 - a_2/7 - 3a_3/5 - 3a_4/5$.

Here, a counterexample to the Wilhelmsen hypothesis can be formulated, which is different from that discussed by L. Bialas-Ciez and P. Goetgheluck [4], to support the claim that the approach leads to a better approximation. Counterexample: Consider the

following polynomial $P(x_1, x_2) = a_0x_1^2 + a_1x_2^2 + a_2x_1x_2 + a_3x_1 + a_4x_2 + a_5$, where $a_5 = -3a_0/7 - 3a_1/7 - a_2/7 - 3a_3/5 - 3a_5/5$. Putting $a_0 = 6, a_1 = 3/4, a_2 = 33/4, a_3 = -6, a_4 = -57/28, a_5 = 3/4, P(x_1, x_2) = 6x_1^2 + 0.75x_2^2 + 8.25x_1x_2 - 6x_1 - 2.04x_2 + 0.75$ is obtained. It is easy to calculate that $\|P\|_{S_2}$ is reached for point $(13/14, 1/14)$, and its value is $297/392$. Similarly, $\|gradP(0, 1)\|_{S_2} = \sqrt{58500}/28$. Hence, this simple and far from optimal approximation value is greater than those of L. Bialas-Ciez and P. Goetgheluck [4]: $\|gradP\|_{S_2}/(4\|P\|_{S_2}) = 2.850293143$.

4. Optimization Problem in a Markov-Type Inequality

The estimation of the constant C in Markov’s inequality can be improved by considering a minimal polynomial and an equilibrium measure. The exemplary calculations of minimal polynomials and their corresponding Markov inequalities on the simplex are shown in Table 2. In the first example of the polynomial, the obtained constant is equal to the value reported by L. Bialas-Ciez and P. Goetgheluck [4]. In the remaining two polynomials, the constants are improved. However, a slight change in the coefficients decreases the constant, as is shown in the fourth example. Therefore, in this paper, the polynomial is proposed as a result of a nontrivial optimization problem in which an improvement of the estimate of the constant C is considered.

Table 2. Constants C in Markov inequalities on the simplex for exemplary polynomials.

Example	Minimal Polynomial $P(x, y)$	$\ P\ _{S_2}$	$\ gradP\ _{S_2}$	$\ gradP\ _{S_2}/4\ P\ _{S_2}$
1	$y^2 - y + \frac{1}{9} + xy - \frac{4}{9}x^2 + \frac{2}{9}x$	5/36	$\sqrt{202}/9$	$C = 2.842534081$
2	$y^2 - y + \frac{1}{8} + \frac{1}{8}x^2 + \frac{33}{24}xy - \frac{19}{56}$	99/784	$10\sqrt{585}/168$	$C = 2.850293143$
3	$y^2 - y + \frac{1}{8} + \frac{61}{40}xy + \frac{1}{4}x^2 - \frac{1}{2}x$	221/1760	$\sqrt{3281}/40$	$C = 2.851041419$
4	$y^2 - y + \frac{1}{8} + \frac{3}{2}xy + \frac{1}{4}x^2 - \frac{1}{2}x$	1/8	$\sqrt{2}$	$C = 2.828427125$

In this paper, the constant in Markov’s inequality on the simplex for the proposed minimal polynomials is estimated. Assuming that

$$\|gradP(x, u)\|_{S_2}/(4\|P(x, u)\|_{S_2}) = C(x, u), \tag{2}$$

where P is one of the following minimal polynomials:

$$P_1 = x_2^2 - a_0x_2 + a_1 + a_2x_1x_2 + a_3x_1^2 + a_4x_1, \tag{3}$$

$$P_2 = x_1^2 - a_0x_1 + a_1 + a_2x_1x_2 + a_3x_2^2 + a_4x_2, \tag{4}$$

$$P_3 = x_1^2 - x_1 + 1/8 + a_0x_1x_2 + a_1x_2^2 + a_2x_2, \tag{5}$$

$$P_4 = x_2^2 - x_2 + 1/8 + a_0x_1x_2 + a_1x_1^2 + a_2x_1, \tag{6}$$

and $u = [a_0, a_1, a_2, a_3, a_4]$, in order to determine the greatest value of the function $C(x, u)$, an optimization problem is formulated.

R. M. Aron and M. Klimek [27] gave explicit formulas for supremum norms: $\|(a, b, c)\|_{\mathbb{R}} = \sup\{|ax^2 + bx + c| : x \in [-1, 1]\}$ and $\|(a, b, c)\|_{\mathbb{C}} = \sup\{|az^2 + bz + c| : z \in \mathbb{C}, |z| \leq 1\}$, where $(a, b, c) \in \mathbb{R}^3$.

The vectors $u = [a_0, a_1, a_2, a_3, a_4]$ and $x = [x_1, x_2]$ are regarded as decision variables in the optimization problem. Consequently, \hat{u} and \hat{x} denote the vectors that maximize the $C(x, u)$. Considering the relationship between P_1 and P_2 , or P_3 and P_4 , only the coefficients that maximize their corresponding $C_1(x, u)$ and $C_3(x, u)$ need to be found, i.e., they are then applied to the remaining polynomials. For polynomials P_3 and $P_4, a_3 = a_4 = 0$. The calculation of Equation (2) requires finding the extrema of the polynomial P inside the simplex limited by values $(x_1, x_2) \in \{(0, 0), (0, 1), (1, 0)\}$. Therefore, the following bilevel optimization problem is solved. Here, one optimization task is nested within the other, i.e., the decision variables of the upper problem are the parameters in the lower one, and the

lower problem is a constraint to the upper problem, limiting the feasible u values to those that are lower-level optimal [28].

The problem is given by

$$\begin{aligned} & \max_{u \in \mathbb{R}^3, x \in [0,1]} C(x, u) \\ & \text{subject to } \hat{x} = \operatorname{argmax}_{x \in [0,1]} \{C(x, u) : x_1 + x_2 - 1 \leq 0\}, \end{aligned} \tag{7}$$

where *argmax* denotes the arguments of the function that maximize its value.

Finally, the values of \hat{x} and \hat{u} are found as the solutions to the optimization problem. In order to determine the greatest value of C , the nonlinearity of the considered objective function, as well as the existence of multiple local optima, must be taken into account. Here, different values of u can lead to the same value of C , or close values of u can result in a large difference between the corresponding values of C . This can be seen in Figure 1, in which only a small part of the search space is visualized. For the visualization, C is calculated, keeping one coefficient constant and selecting the remaining one from the range $[-1, 1]$ with the step 0.1. The coefficients are selected from the set $\{-0.1, 0, 0.1\}$. The difference between the values of the constant is purposely small to show the changes of values of C . Each subplot in Figure 1 contains a marker to additionally highlight such changes.

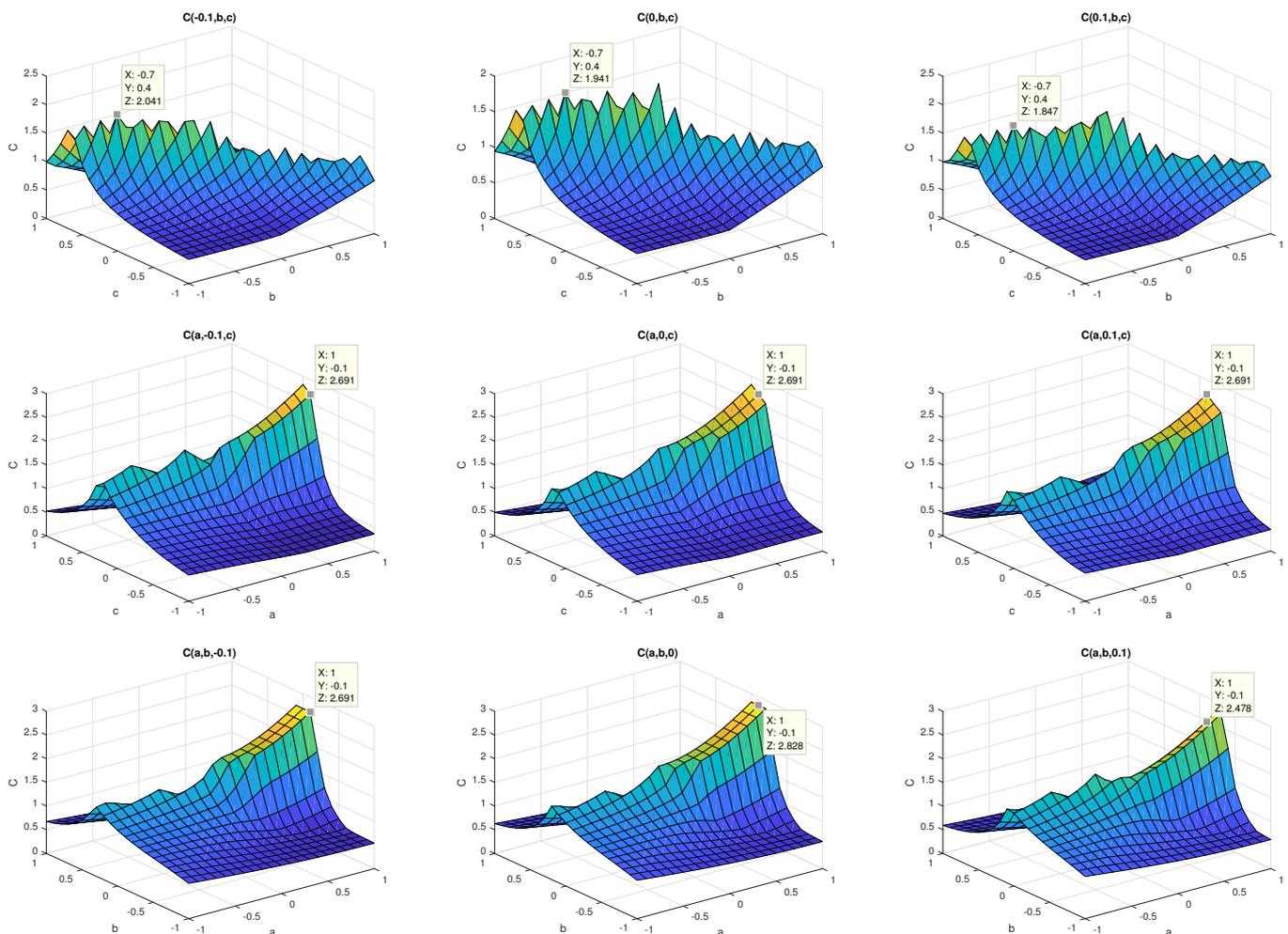


Figure 1. Influence of the coefficients u on the objective function C_4 . Three coefficients are selected: a_0 (first row), a_1 (second row), and a_2 (third row).

To solve the defined optimization problem, a suitable method should be determined. Therefore, in this paper, a hybrid approach is proposed. It is justified by the need to find the extrema of the polynomial P inside the simplex for the computation of C . Such extrema are found using the Interior Point Algorithm (IPT) [29,30], i.e., \hat{x} is found for an examined u from the upper level. The upper-level optimization is more difficult. Since the considered problem has not been formulated and solved in the literature, it cannot be determined in advance which method should be used to determine the solution. Therefore, taking into account possible time-consuming computations, three popular meta-heuristics are used and compared: the Particle Swarm Optimization method (PSO) [31–33], Genetic Algorithm (GA) [34,35], and Simulated Annealing technique (SA) [36,37]. The application of population-based algorithms for solving different bilevel problems is popular (e.g., [28]).

The PSO is a global evolutionary method based on swarm intelligence. It finds the optimal solution by exploring the search space using a set of particles. Each particle represents a candidate solution to the problem. The PSO mimics the movement of a bird flock or a fish school; hence, the dynamics of the set of particles in the solution space are iteratively governed by the following equations [31]:

$$p_i(n) = p_i(n-1) + v_i(n)$$

$$v_i(n) = wv_i(n-1) + c_1 \text{rnd}_1(pb_i - p_i(n-1)) + c_2 \text{rnd}_2(pB - p_i(n-1)),$$

where $p_i(n)$ and $v_i(n)$ are the position and the velocity of the i -th particle at time n , $n = 1, 2, \dots, N$; pb_i is the best position of the i -th particle, while pB is the best position of the swarm at n . w is the inertia weight, and c_1 and c_2 are coefficients constraining the parts of the equation that simulate cognitive and social factors of the swarm [31–33]; rnd_1 and rnd_2 are random numbers drawn from the range $[0, 1]$. The termination criterion of the algorithm is the number of iterations N or the violation of the minimal acceptable difference in the objective function C between consecutive iterations. Both criteria are commonly used in most heuristic optimization algorithms. Finally, the PSO returns $\hat{u} = pB(N)$.

The GA uses a population of individuals (called chromosomes), representing candidate solutions [34,35]. In the GA, at each generation (iteration), a population of individuals is assessed to select the best solutions, which are then used to create the population in the next generation. The assessment is based on the values of the objective function, while new individuals are obtained using crossover and mutation operators. The crossover combines the information carried by two selected parents in one solution by mixing values of their vector representations. The mutation, in turn, introduces random changes to the individual. After the predefined number of generations, the algorithm returns the best individual \hat{u} .

The simulated annealing, unlike the population-based PSO and GA, uses much fewer resources and requires fewer objective function computations, since it processes only one solution at each iteration [36,37]. The SA mimics the process of heating a metal and slowly lowering its temperature. The algorithm starts from a random solution, which is iteratively modified. At each iteration, reflecting a temperature drop, the algorithm modifies the current solution and compares the values of the objective function before and after the modification. If a new solution is better or the difference between their values is within a predefined range, the modification is accepted. The acceptance of a slightly worse solution helps the method to escape local minima [36,37].

All of these methods are commonly used for solving nonlinear optimization problems with local optima. However, their suitability for a given problem is mostly determined experimentally. In this paper, they are used to determine \hat{u} . However, in order to calculate C for any u , the lower-level optimum \hat{x} must be solved each time. In this paper, it is obtained using the IPT algorithm [29,30]. Finally, the considered problem is solved using the hybrid approach [28]. In the IPT algorithm, an optimization problem is defined as

$$\min_x f(x)$$

$$\text{subject to } h(x) = 0 \quad \text{and} \quad g(x) \leq 0, \quad (8)$$

where $f(x)$ is objective function to be minimized and h and g are equality and inequality constraints, respectively. The solved approximate problem is defined as

$$\begin{aligned} \min_{x,s} f_{\mu}(x,s) &= \min_{x,s} f(x) - \mu \sum_i \ln(s_i) \\ \text{subject to } h(x) &= 0 \quad \text{and} \quad g(x) + s = 0, \end{aligned} \tag{9}$$

where s_i are slack variables and $\mu > 0$ is a barrier parameter. As μ converges to zero, the sequence of approximate solutions converges to the minimum of f [29,30].

In the numerical experiments, the solution processed by heuristic algorithms was represented by a five-dimensional vector of floating-point values, since $u = [a_0, a_1, a_2, a_3, a_4]$, and the IPT used a two-dimensional vector $x = [x_1, x_2]$. To ensure a fair comparison of meta-heuristics, each technique was run 50 times, and a maximum of 25,000 objective function calls were performed in each run. The number of function calls was achieved by selecting a large-enough population size, aiming at the best performance of the PSO and GA. The implementations of the algorithms that can be found in the Matlab Optimization Toolbox are used. Algorithms were run with their default parameters: the PSO ($w=[0.1, 1.1]$, $c_1 = 1.4$, $c_2 = 1.4$), the GA ($EliteCount = 0.05$ of the population size, $CrossoverFraction = 0.8$, the scattered crossover, Gaussian mutation, and stochastic uniform selection), and the SA ($InitialTemperature = 100$, temperature is equal to $InitialTemperature * 0.95^{iteration}$). Note that the values in u were not bounded (i.e., $u \in \mathbb{R}^5$), and the GA and PSO, unlike the SA, did not require the specification of the starting point of the optimization, as the initial populations of the solutions were selected randomly. For this reason, the SA started from $u = [0, 0, 0, 0, 0]$. As demonstrated in Table 3, the PSO was able to determine the best C for each polynomial. The algorithms explored the solution space in different ways, obtaining different results.

Table 3. Optimization results obtained by the compared algorithms. The best values of C obtained in the first experiment are written in bold.

Algorithm	C ₁ , C ₂	C ₃ , C ₄
GA	2.861879141	2.852924202
SA	2.917638439	2.854084285
PSO	2.925401560	2.854206134

To further show the difficulty of the optimization task for the used algorithms, Figure 2 presents the convergence of the algorithms towards the optimum. Here, the progress of the set of solutions at each calculation of the objective function is reported. As presented, the approaches with the PSO and GA quickly converge, and most of the time is spent on the exploration of the solution space near the optimum. The SA, unlike the population-based methods, processes one solution at each iteration, which influences its convergence rate. The results reveal that the PSO is better suited to the problem than the GA, since it can determine a better C (see Table 3), and it finds it with fewer function calls. Interestingly, the set of solutions in the first function call is more diverse for the PSO, which may contribute to its superior performance, as in the remaining computation time, it is more focused on escaping the local optima. Similarly, it can be seen that the solutions emerging in the run of the SA exhibit large variability. Note that the number of runs (50) and the maximal number of function calls in these experiments (25,000) are purposely greater than could be concluded from the convergence plots to give the methods enough resources to find better solutions. However, the methods often terminate the computations if the change in the best function value is too small.

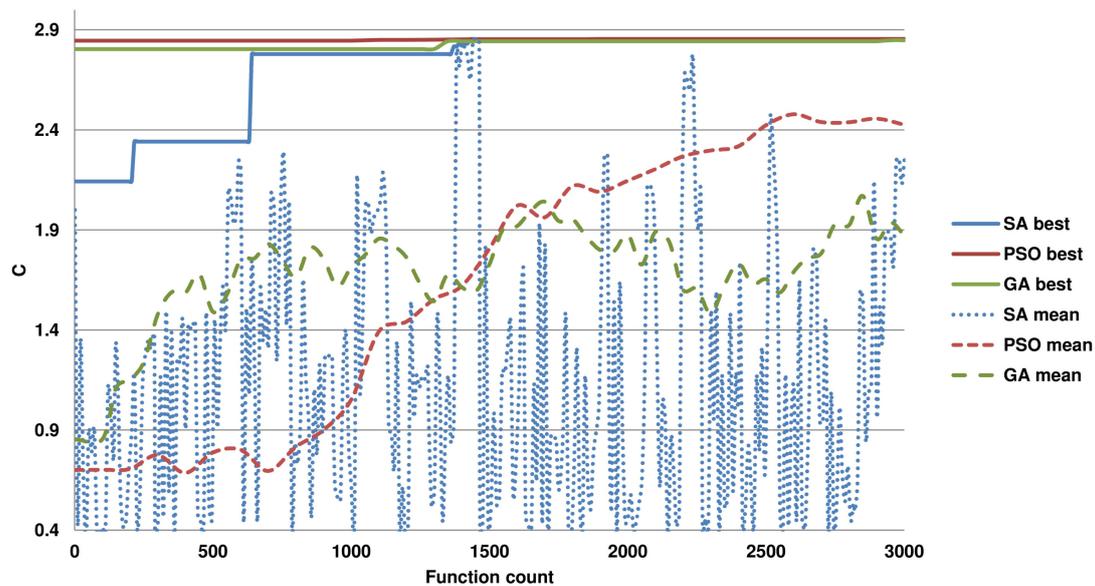


Figure 2. Convergence plots for the values of C_4 for sample runs of the hybrid approach with the Particle Swarm Optimization (PSO), Simulated Annealing (SA), and Genetic Algorithm (GA). The plots contain the best result at each function call and the mean value of C_4 for the processed solutions.

This experiment indicates that the proposed approach with the PSO can provide the best C . Therefore, in the second experiment, the parameters of this algorithm were set to conduct an even better search of the solution space. By increasing the number of particles and generations, the maximal number of function calls was set to 25 million, and the method was run 100 times. Note that the estimated time to finish one run of the PSO with 25 million function calls in Matlab on the i7-6700 K 4 GHz with 64 GB RAM CPU is 335 h. The calculations were run in a parallel manner, and the methods were allowed to terminate the optimization if a function value change was small. They are presented in Table 4. The second experiment improved the already obtained values for C_1 (or C_2) and C_3 (or C_4) at the fifth and eighth decimal places, respectively (see Table 3). The run in which the best C_1 was obtained required 720 thousand function calls, and C_3 was found in a run after the objective function was calculated 1570 thousand times.

Table 4. Values of C obtained in the second experiment and their corresponding coefficients.

C_1, C_2	2.925415050
a_0	−2765.096993
a_1	−1383.048545
a_2	−15148.04101
a_3	−10633.59420
a_4	10846.85282
C_3, C_4	2.854206170
a_0	1.169292858
a_1	−0.121651344
a_2	−0.129695921

5. Conclusions

In this paper, it was demonstrated that the best constant C in Markov’s inequality for the minimal polynomials on the simplex can be determined as a solution to a nested optimization problem. First, the examples of the polynomials that allow estimation of the constant were proposed. Then, the optimization problem was defined and its solution was investigated. Since it could not be determined in advance which heuristic optimization

algorithm should be used, three popular meta-heuristics were adopted and compared. Finally, the optimal value of the constant was obtained with a hybrid of the ITP and PSO.

Future work will focus on the application of the polynomials of the higher degrees to the considered problem. In addition, the usage of the introduced bilevel optimization problem as an evaluation tool for the comparison of the performance of meta-heuristics seems to be a promising future direction.

Author Contributions: Conceptualization, G.S. and M.O.; methodology, G.S. and M.O.; software, M.O.; validation, G.S. and M.O.; formal analysis, G.S.; investigation, G.S. and M.O.; writing—original draft preparation, G.S. and M.O.; writing—revised draft preparation, G.S. and M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Baran, M.; Bialas-Ciez, L. On the behaviour of constants in some polynomial inequalities. *Ann. Pol. Math.* **2019**, *123*, 43–60. [CrossRef]
2. Baran, M.; Bialas-Ciez, L.; Milowka, B. On the best exponent in Markov's inequality. *Potential Anal.* **2013**, *38*, 635–651. [CrossRef]
3. Baran, M. Markov inequality on sets with polynomial parametrization. *Ann. Polon. Math.* **1994**, *60*, 69–79. [CrossRef]
4. Bialas-Ciez, L.; Goetgheluck, P. Constants in Markov's inequality on convex sets. *East J. Approx.* **1995**, *1*, 379–389.
5. Bun, M.; Thaler, J. Dual lower bounds for approximate degree and Markov-Bernstein inequalities. *Inf. Comput.* **2015**, *243*, 2–25.
6. Nadzhmuddinov, D.; Subbotin, Y.N. Markov Inequalities for Polynomials on Triangles. *Math. Notes Acad. Sci. USSR* **1989**, *46*, 627–631. [CrossRef]
7. Baran, M. Bernstein type theorems for compact sets in \mathbb{R}^N revisited. *J. Approx. Theory* **1994**, *79*, 190–198. [CrossRef]
8. Baran, M. Complex equilibrium measure and Bernstein type theorems for compact sets in \mathbb{R}^n . *Proc. Amer. Math. Soc.* **1995**, *123*, 485–494.
9. Sofonea, F.; Tincu, I. On an Inequality for Legendre Polynomials. *Mathematics* **2020**, *8*; 2044.
10. Ozisik, S.; Riverce, B.; Warburton, T. On the Constants in Inverse Inequalities in L_2 . Technical Report. 2010. Available online: <https://hdl.handle.net/1911/102161> (accessed on 5 December 2020)
11. Piazzon, F.; Vianello, M. A note on total degree polynomial optimization by Chebyshev grids. *Optim. Lett.* **2018**, *12*, 63–71.
12. Piazzon, F.; Vianello, M. Markov inequalities, Dubiner distance, norming meshes and polynomial optimization on convex bodies. *Optim. Lett.* **2019**, *13*, 1325–1343. [CrossRef]
13. Sommariva, A.; Vianello, M. Discrete norming inequalities on sections of sphere, ball and torus. *arXiv* **2018**, arXiv:1802.01711.
14. Wilhelmsen, D.R. A Markov inequality in several dimensions. *J. Approx. Theory* **1974**, *11*, 216–220. [CrossRef]
15. Sarantopoulos, Y. Bounds on the derivatives of polynomials on Banach spaces. *Math. Proc. Camb. Philos. Soc.* **1991**, *110*, 307–312. [CrossRef]
16. Kroó, A. Révész, S. On Bernstein and Markov-type inequalities for multivariate polynomials on convex bodies. *J. Approx. Theory* **1999**, *99*, 134–152. [CrossRef]
17. Kroó, A. On the existence of optimal meshes in every convex domain on the plane. *J. Approx. Theory* **2019**, *238*, 26–37. [CrossRef]
18. Davydov, O. *Smooth Finite Elements and Stable Splitting*; Berichte "Reihe Mathe-matik" der Philipps-Universit at Marburg: Marburg, Germany, 2007.
19. Angermann, L.; Heneke, C. Interpolation, Projection and Hierarchical Bases in Discontinuous Galerkin Methods. *Numer. Math. Theory Methods Appl.* **2015**, *8*, 425–450.
20. Baran, M.; Kowalska, A.; Ozorka, P. Optimal factors in Vladimir Markov's inequality in L^2 Norm. *Sci. Tech. Innov.* **2018**, *2*, 64–73.
21. Bialas-Ciez, L.; Jedrzejowski, M. Transfinite Diameter of Bernstein Sets in \mathbb{C}^N . *J. Inequal. Appl.* **2002**, *7*, 393–404.
22. Bloom, T.; Calvi, J.P. On multivariate minimal polynomials. *Math. Proc. Camb. Phil. Soc.* **2000**, *129*, 417–432.
23. Bialas-Ciez, L.; Calvi, J.P. Homogeneous minimal polynomials with prescribed interpolation conditions. *Proc. Am. Math. Soc.* **2016**, *368*, 8383–8402. [CrossRef]
24. Newman, D.J.; Xu, Y. Tchebycheff polynomials on a triangular region. *Constr. Approx* **1993**, *9*, 543–546. [CrossRef]
25. Peherstorfer, F. Minimal polynomials for compact sets of the complex plane. *Constr. Approx* **1996**, *12*, 481–488. [CrossRef]
26. Davis, P.J. *Approximation and Interpolation*; Dover Publication: New York, NY, USA, 1975.
27. Aron, R.M.; Klimek, M. Supremum norms for quadratic polynomials. *Arch. Math.* **2001**, *76*, 73–80. [CrossRef]
28. Sinha, A.; Malo, P.; Deb, K. Evolutionary algorithm for bilevel optimization using approximations of the lower level optimal solution mapping. *Eur. J. Oper. Res.* **2017**, *257*, 395–411. [CrossRef]

29. Byrd, R.H.; Gilbert, J.C.; Nocedal, J. A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming. *Math. Prog.* **2000**, *89*, 149–185. [[CrossRef](#)]
30. Byrd, R.H.; Hribar, M.E.; Nocedal, J. An Interior Point Algorithm for Large-Scale Nonlinear Programming. *SIAM J. Optim.* **1999**, *9*, 877–900. [[CrossRef](#)]
31. Gao, Y.; Zhang, G.; Lu, J.; Wee, H.-M. Particle swarm optimization for bi-level pricing problems in supply chains. *J. Glob. Optim.* **2011**, *51*, 245–254. [[CrossRef](#)]
32. Pedersen, M.E. *Good Parameters for Particle Swarm Optimization*; Hvas Laboratories: Luxembourg, 2010.
33. Zhang, Y.; Wang, S.; Ji, G. A Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications. *Math. Probl. Eng.* **2015**, *2015*, 38.
34. Goldberg, D.E. Genetic Algorithms in Search. In *Optimization & Machine Learning*; Addison-Wesley: Boston, MA, USA, 1989.
35. Yang, J.; Zhang, M.; He, B.; Yang, C. Bi-level programming model and hybrid genetic algorithm for flow interception problem with customer choice. *Comput. Math. Appl.* **2009**, *57*, 1985–1994. [[CrossRef](#)]
36. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–680. [[CrossRef](#)]
37. Sahin, K.H.; Ciric, A.R. A dual temperature simulated annealing approach for solving bilevel programming problems. *Comput. Chem. Eng.* **1998**, *23*, 11–25. [[CrossRef](#)]