A Mechanical Method for Isolating Locally Optimal Points of Certain Radical Functions

Zhenbing Zeng[1(✉)], Yaochen Xu[1,2], Yu Chen[1], and Zhengfeng Yang[3]

[1] Department of Mathematics, Shanghai University, Shanghai 200444, China
zbzeng@shu.edu.cn

[2] Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Science, Shanghai 200031, China
xuyaochen@sibcb.ac.cn

[3] School of Software Engineering, East China Normal University, Shanghai 200062, China
zfyang@sei.ecnu.edu.cn

Abstract. In this paper, we present a symbolic computation method for constructing a small neighborhood U around a known local optimal maximal or minimal point x_0 of a given smooth function $f : \mathbb{R}^n \to \mathbb{R}$ that contains radical or rational expressions of several variables, so that x_0 is also the global optimal point of $f(x)$ restricted to the small neighborhood U. The constructed small neighborhood can be used to prove that $f(x_0)$ is the global optimum of f in a rather large region M with $U \subset M$ via exact numeric computation like interval evaluation and branch-and-bound technology.

Keywords: Locally optimal points · Isolating algorithm · Radical function · Symbolic computation

1 Introduction

In some geometric optimization problems, we want to calculate the maximal value of a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$ over some domain $M \subset \mathbb{R}^n$ which contains radical (or rational, trigonometrical) expressions. Usually, the objective function f is smooth, i.e., it has continuous derivatives up to any desired order over M. Therefore, applying numerical experiments the *de facto* optimal point of f can be observed with very big confidence, and it is also relatively easy to verify that the optimal point x_0 obtained from numerical searching is actually a local optimal point, namely, the partial derivatives of f with respect to each variable is zero at x_0, and the Hessian matrix of f at x_0 is positive-(semi-) definite or negative-(semi-)definite. In many cases, the numerical computation also shows that x_0 is the unique local optimal point of the objective function,

Supported by National Natural Science Foundation of China (12171159, 12071282).

F. Boulier et al. (Eds.): CASC 2022, LNCS 13366, pp. 377–396, 2022.
https://doi.org/10.1007/978-3-031-14788-3_21

but a strict mathematical proof is hard due to intermediate expression swell in symbolic computation.

For example, let $P_i = (x_i, y_i, z_i)$ $(i = 1, 2, \ldots, 6)$ be six points on the unit sphere S^2 and suppose we want to find the maximum of the sum of their pairwise Euclidean distances, $d = \sum_{1 \le i < j \le 6} ||P_i - P_j||_2$, where

$$||P_i - P_j||_2 = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}.$$

To avoid the manifold solution of this optimal problem generated by the rigid movements on S^2, we may assume that one point has been fixed at the North Pole, and another point has been fixed on the prime meridian. Then, both of Monte Carlo search and the grid search (see, e.g., [1,9,10]) show that the maximum of d is $22.9705\ldots \approx 6+12\sqrt{2}$, and the local optimal points are the following unique ones:

$$(0, 0, 1),\ (0, 0, -1),\ (1, 0, 0),\ (-1, 0, 0),\ (0, 1, 0),\ (0, -1, 0).$$

To the best of our knowledge, no mathematical proof has been given to this conjecture yet.

Generally, if $x_0 \in M$ is the unique local maximal point of a continuous function $f(x)$ formed by finitely many steps of the four basic arithmetic operations, the radical, the exponential, and the trigonometrical functions of n variables $x_1, x_2, \ldots, x_n$ on a compact domain $M \subset \mathbb{R}^n$, then, by interval evaluation of $f(x)$, we can construct a neighborhood

$$U(x) = [x_1 - \varepsilon, x_1 + \varepsilon] \times [x_2 - \varepsilon, x_2 + \varepsilon] \times \cdots \times [x_n - \varepsilon, x_n + \varepsilon] \subset M,$$

for any point $x = (x_1, x_2, \ldots, x_n) \in M \setminus \{x_0\}$, where $\varepsilon = \varepsilon(x) > 0$ is dependent on x, so that the upper bound of $f(x)$ on $U(x)$ is less than $f(x_0)$. If we can also find a neighborhood $U_0 = U(x_0)$ of x_0 so that restricted on $U(x_0)$, $f(x) \le f(x_0)$, then we will get a family of neighborhood $\{U(x)|x \in M\}$ that covers the set M. According to the compactness of M, we would find a finite subset $\{U_1, U_2, \ldots, U_N\}$ of the family that satisfies

$$M \subset U(x_0) \cup U_1 \cup U_2 \cup \cdots \cup U_N,$$

and on each U_i, $f(x) \le f(x_0)$. Clearly, if we could generate all neighborhoods $U(x)$ for every point $x \in M \setminus \{x_0\}$ in advance, then we would be able to produce a proof to the original optimization problem. To utilize this idea on computer for a machine proof, we may implement this through the following two procedures:

`Procedure 1: isolate the local optimal point.` Construct a function $g(x)$ which has x_0 as the unique maximal point with $g(x_0) = f(x_0)$, and a neighborhood U_0 of x_0 that satisfies $f(x) \le g(x)$ for $x \in U_0$, and, therefore, $f(x) \le f(x_0)$ on U_0.

`Procedure 2: ‘‘divide-and-conquer’’ outside the isolated regions.` Partition $M \setminus U_0$ into a sequence of cubes $D_1, D_2, \ldots, D_m$ in $\mathbb{R}^n$ where D_i, D_j have no common interior for $1 \le i < j \le m$, and apply the interval evaluation (or grid interpolation) to estimate the upper bound $u(D_i)$ of f on each cube $D_i (1 \le i \le m)$. If $u(D_i) \ge f(x_0)$ for some $i (1 \le i \le m)$, then divide D_i into 2^n smaller cubes $D_{i,j}$ $(j = 1, 2, \dots, 2^n)$ in $\mathbb{R}^n$ whose edge length is one half of that of D_i, and estimate the upper bounds $u(D_{i,j})$ of $f(x)$ on the newly obtained cubes $D_{i,j}$. Recursively do this until the upper bound of every cube D_I produced in this process satisfies $u(D_I) < f(x_0)$. This process will be terminated after finitely many steps of subdivision provided $\sup_{x \in M \setminus U_0} f(x) < f(x_0)$, since according to Taylor's theorem, we have

$$\begin{aligned} f(x) &= f(x^c_{D_I}) + (x - x^c_{D_I}) \nabla f(t x^c_{D_I} + (1-t)x) \\ &\le f(x^c_{D_I}) + \frac{\sqrt{n}}{2} \mathrm{edge}(D_I) \cdot B_0, \end{aligned} \tag{1}$$

for all $x \in D_I$. Here $x^c_{D_I}$ is the barycenter of D_I, $t = t(x) \in [0, 1]$, $\mathrm{edge}(D_I)$ is the edge length of D_I, and B_0 is the following constant:

$$B_0 = \sup_{x \in M \setminus U_0} \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 + \cdots + \left(\frac{\partial f}{\partial x_n}\right)^2} < +\infty,$$

and we may assume that the estimated upper bound $u(D_I)$ of $f(x)$ on every cube D_I satisfies the following inequality

$$u(D_I) \le f(x^c_{D_I}) + \frac{\sqrt{n}}{2} \mathrm{edge}(D_I) \cdot B_0. \tag{2}$$

Therefore, if the subdivision cannot be completed in finite steps, we would get a sequence $D_i, D_{i,j_1}, D_{i,j_1,j_2}, \dots, D_{i,j_1,j_2,\dots,j_k}$ $(1 \le i \le m, 1 \le j_k \le 2^n, k = 1, 2, \dots)$ that satisfies $u(D_{I_k}) \ge f(x_0)$ for $D_{I_k} = D_{i,j_1,j_2,\cdots,j_k}$ $(k = 1, 2, \cdots)$, which leads to

$$\lim_{k \to \infty} f(x^c_{D_k}) \ge f(x_0),$$

and contradicts the assumption $\sup\limits_{x \in M \setminus U_0} f(x) < f(x_0)$.

To our knowledge, this approach to automated proof of inequalities was suggested by Jingzhong Zhang in the late 1980s for proving an inequality of Zirakzadeh (see [14] and [2]). A detailed description of Zhang's method can be found in [13] in Chinese. Later the method was used in [5] and [12] for proving two other geometric inequalities related to optimal distribution of points on sphere and hemisphere. However, the technique of `Procedure 1` is not described in a general term in these case studies, so it is still difficult to apply the new method to process other unsolved or complicated problems directly.

This paper is aiming to give a general symbolic algorithm of `Procedure 1` for a class of smooth functions formed by a sum of several radical expressions. Namely, assume that $f : \mathbb{R}^n \to \mathbb{R}$ has the following form:

$$f = c_1 \sqrt{g_1(x_1, x_2, \dots, x_n)} + \cdots + c_k \sqrt{g_k(x_1, x_2, \dots, x_n)},$$

where $c_1, \ldots, c_k$ are real numbers and $g_j(x_1, x_2, \ldots, x_n)\,(j = 1, \ldots, k)$ are polynomials or rational functions of polynomials, and the point $x_0 \in \mathbb{R}^n$ satisfies the conditions

$$\frac{\partial f}{\partial x_i}(x_0) = 0,\ i = 1, 2, \ldots, n;$$

and

$$H_0 := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} (x_0)$$

is negative-semi-definite. We explain how to construct a quadratic form $q(x) = q(x_1, x_2, \ldots, x_n)$ and a neighborhood $U_0 \subset \mathbb{R}^n$ of x_0 so that

(1) $q(x_0) = f(x_0)$, x_0 is the unique maximal point of $q(x)$, and
(2) $f(x) \leq q(x)$ for $x \in U_0$.

Note that the methods of local analysis in [5,12,13] are implemented for triangular functions and some special radical functions. We shall present our algorithm in a more general form. Actually, our algorithm gives a constructive approach to a special case (for $k = 0$ or n) of the Morse Lemma (see, e.g., [3,7]), which asserts that if $f : \mathbb{R}^n \to \mathbb{R}$ is a function of class C^∞ for which $x_0 = 0$ is a non-degenerate critical point, namely $\nabla f(0) = 0$ and the Hessian at x_0 has trivial kernel, then in *some neighbourhood* U of x_0 there is a local C^∞ coordinate system, namely a C^∞ diffeomorphism $\varphi : U \to V \subset \mathbb{R}^n$ with $\varphi(0) = 0$ and such that $\tilde{f} = f \circ \varphi^{-1}$ takes the form

$$\tilde{f}(x) = f(0) - x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2.$$

Several quantitative forms of the Morse Lemma can be found in [4,6,8,11], yet a symbolic computation method cannot be directly derived from the literature.

The paper is organized as follows. In Sect. 2, we show how to find quadratic bounds of an algebraic surface in the neighborhoods of a critical point; in Sect. 3, we extend the method to rational and certain radical functions. In Sect. 4, we shall apply the method to do local critical analysis for the spherical six-point problem. The `Maple` computation in this paper is implemented on Maple version 18.00.

2 Quadratic Local Upper Bound of Polynomials

The following analytic definition of local optimal (extremum, maximal or minimal) of a real-valued function can be found in any calculus text book.

Definition 1. *A real-valued function f defined on a real-line is said to have a local (or relative) maximum point at the point x_0, if there exists some $\varepsilon > 0$ such that $f(x) \leq f(x_0)$ when $|x - x_0| < \varepsilon$. The value of the function at this point is called maximum of the function. Similarly, a function has a local minimum point at x_0, if $f(x) \geq f(x_0)$ when $|x - x_0| < \varepsilon$. The value of the function at this point is called minimum of the function.*

For functions of several variables, a neighborhood $U(x_0, \varepsilon)$ of the point x_0 is used to substitute the interval $|x - x_0| < \varepsilon$. It is well known that the local extrema can be found by Fermat's theorem, which states that they must occur at critical points (also called stationary points).

Theorem 1 (Fermat's theorem). *Let $f : (a, b) \to R$ be a function and suppose that $x_0 \in (a, b)$ is a local maximum of f. If f is differentiable, then $f'(x_0) = 0$. And exactly the same statement is true in higher dimensions.*

One can distinguish whether a critical point is a local maximum or local minimum by using the second derivative test. In calculus, the *second derivative test* is a criterion for determining whether a given critical point of a function is a local maximum or a local minimum using the value of the second derivative at the point. The test states: if the function f is twice differentiable at a stationary point x_0, then

- If $f''(x_0) < 0$ then f has a local maximum at x_0.
- If $f''(x_0) > 0$ then f has a local minimum at x_0.
- If $f''(x_0) = 0$, the second derivative test says nothing about the point x_0.

For a function of more than one variable, the second derivative test generalizes to a test based on the eigenvalues of the function's Hessian matrix at the stationary point. In particular, assuming that all second order partial derivatives of f are continuous in a neighbourhood of a stationary point x_0, and the eigenvalues of the Hessian at x_0 are all positive, then x_0 is a local minimum. If the eigenvalues are all negative, then x_0 is a local maximum, and if some are positive and others are negative, then the point x_0 is a saddle point. If the Hessian matrix is singular, then the second derivative test is inconclusive. Note that the second derivative test concludes only the existence of a neighbourhood U_0 of x_0, where the function f satisfies $f(x) \geq f(x_0)$, or $f(x) \leq f(x_0)$, for all points $x \in U_0$.

It is easy to see that for a quadratic polynomial $p(x)$ with n-variables, if $x_0 = (0, 0, \ldots, 0) \in \mathbb{R}^n$ is a local maximum point, then

$$p(x) = p_0 + \frac{1}{2}(x_1, x_2, \ldots, x_n)H_0(x_1, x_2, \ldots, x_n)^T,$$

where $p_0 = p(0, 0, \ldots, 0)$ and H_0 is a negative-semi-definite symmetric matrix, so under certain orthogonal transform of Cartesian coordinates

$$(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n) \cdot P.$$

We may express the polynomial p using the new coordinates as

$$p(y_1, y_2, \ldots, y_n) = p_0 + \frac{1}{2}(\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2), \tag{3}$$

where P is an orthogonal matrix and $\lambda_i \leq 0$ $(i = 1, 2, \ldots, n)$ are the eigenvalues of H_0, which also shows that $x = 0$ is the global optimal of $p(x)$. In geometry, this shows that in $\mathbb{R}^{n+1}$, the algebraic surface

$$F := \{(x_1, x_2, \ldots, x_n, z) | z - p(x_1, x_2, \ldots, x_n)\} = 0$$

lies at one side of the tangent space $T_0F : z = 0$ of F at 0.

For polynomial $p(x)$ of degree $d \geq 3$, if $x = 0$ is a local maximum of p, then there is a neighborhood U_0 of $0 \in \mathbb{R}^n$ so that in the local region $U_0 \times \mathbb{R} \subset \mathbb{R}^{n+1}$, the surface $F : z - p(x) = 0$ and the tangent space T_0F can be separated by a quadratic surface $F_1 : z - q(x) = 0$, where $q(x)$ is a quadratic polynomial which has $x = 0$ as its maximal point, and therefore, the algebraic surface $z - p(x) = 0$ lies under its tangent space at 0. We will show that the quadratic polynomial can be constructed using symbolic computation. Namely, we have the following result.

Theorem 2. *Assume that $p(x) = p(x_1, x_2, \ldots, x_n)$ is a polynomial of degree $d \geq 3$ and $x = (0, 0, \ldots, 0)$ is a local maximum of $p(x)$ satisfying the condition that the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the Hessian matrix of $p(x)$ at $x = 0$ are all negative. Then we can construct a neighborhood U_0 of $x = 0$ and a quadratic polynomial $q(x_1, x_2, \ldots, x_n)$ satisfies*

(i) $q(0) = p(0)$,
(ii) $\frac{\partial q}{\partial x_i} = 0$ *for* $i = 1, 2, \ldots, n$,
(iii) the Hessian matrix $H_0(q)$ *is negative-definite, and*
(iv) $p(x) \leq q(x)$ *for all* $x \in U_0$.

We may call $q(x)$ in Theorem 2 a quadratic local upper-bound of polynomial $p(x)$. In order to prove this theorem, we need to consider the degree-j homogeneous part of polynomial $p(x)$ for each degree $j \geq 3$. We have the following lemma.

Lemma 1. *For any integer $j \geq 3$ and homogeneous polynomial*

$$h_j(x_1, x_2, \ldots, x_n) = \sum_{\substack{d_1, d_2, \ldots, d_n \geq 0 \\ d_1 + d_2 + \cdots + d_n = j}} c_{d_1, d_2, \ldots, d_n} x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n},$$

with real coefficients, then there exists constant numbers $k_1, k_2, \ldots, k_n \geq 0$, such that for any positive number N and for real numbers $x_1, x_2, \ldots, x_n \in (-1/N, 1/N)$, the inequality

$$|h_j(x_1, x_2, \ldots, x_n)| \leq \frac{1}{jN^{j-2}}(k_1 x_1^2 + k_2 x_2^2 + \cdots + k_n x_n^2)$$

holds.

Proof. For any j, real numbers $z_1, z_2, \ldots, z_j \in (-1/N, 1/N)$ and any combination (k, l) of $1, 2, \ldots, j$, we have

$$z_1 z_2 \cdots z_j \leq \frac{1}{2N^{j-2}}(z_k^2 + z_l^2). \tag{4}$$

Construct this inequality for all $\binom{j}{2} = j(j-1)/2$ two-member combinations of $1, 2, \ldots, j$, and sum up them to obtain

$$\binom{j}{2} z_1 z_2 \cdots z_j \leq \frac{1}{2N^{j-2}}(j-1)(z_1^2 + z_2^2 + \cdots + z_j^2).$$

Therefore

$$z_1 z_2 \cdots z_j \leq \frac{1}{jN^{j-2}}(z_1^2 + z_2^2 + \cdots + z_j^2),$$

and

$$\begin{aligned} x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n} &= \overbrace{x_1 \cdots x_1}^{d_1} \times \overbrace{x_2 \cdots x_2}^{d_2} \times \cdots \times \overbrace{x_n \cdots x_n}^{d_n} \\ &\leq \frac{1}{jN^{j-2}}(d_1 x_1^2 + d_2 x_2^2 + \cdots + d_n x_n^2), \end{aligned} \tag{5}$$

for any monomial $x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}$ of degree j. Applying inequality (5) to each monomial of the homogeneous polynomial $h_j(x_1, x_2, \ldots, x_n)$, we have

$$\begin{aligned} |h_j| &\leq \sum_{d_1+d_2+\cdots+d_n=j} |c_{d_1,d_2,\ldots,d_n} x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}| \\ &\leq \sum \frac{1}{jN^{j-2}} |c_{d_1,d_2,\ldots,d_n}| (d_1 x_1^2 + d_2 x_2^2 + \cdots + d_n x_n^2) \\ &= \frac{1}{jN^{j-2}} \left(k_1 x_1^2 + k_2 x_2^2 + \cdots + k_n x_n^2 \right). \end{aligned} \tag{6}$$

Here $k_1, k_2, \ldots, k_n$ are positive real numbers defined by

$$k_i = \sum_{d_1+\cdots+d_n=j} d_i |c_{d_1,d_2,\ldots,d_j}|, \quad i = 1, 2, \cdots, n.$$

This completes the proof of Lemma 1. □

Remark 1. Taking $C_j = \max\{k_1, k_2, \ldots, k_n\}$, then inequality (6) can be written in the following simple form:

$$|h_j(x_1, x_2, \ldots, x_n)| \leq \frac{C_j}{jN^{j-2}}(x_1^2 + x_2^2 + \cdots + x_n^2). \tag{7}$$

Now we give a proof of Theorem 2.

Proof. Let $p_0 = p(0)$, H_0 be the Hessian matrix of $p(x)$ at $x = 0$, and $\lambda_1, \lambda_2, \ldots, \lambda_n$ the eigenvalues of H_0. Then $\lambda_i < 0 \, (i = 1, 2, \ldots, n)$ according to the assumption. We can express p as follows:

$$\begin{aligned} p(x) = p_0 &+ \frac{1}{2}(x_1, x_2, \ldots, x_n) H_0 (x_1, x_2, \ldots, x_n)^T \\ &+ H_3(x_1, x_2, \ldots, x_n) + \cdots + H_d(x_1, x_2, \ldots, x_n), \end{aligned} \tag{8}$$

where H_j $(j = 3, \ldots, d)$ are homogeneous polynomials of degree j, respectively.

Applying Lemma 1, for each j $(j = 3, \ldots, d)$ we compute a sequence of constants $k_1^{(j)}, k_2^{(j)}, \ldots, k_n^{(j)}$ that satisfy the following inequality

$$|H_j(x_1, x_2, \ldots, x_n)| \leq \frac{1}{jN^{j-2}}(k_1^{(j)} x_1^2 + k_2^{(j)} x_2^2 + \cdots + k_n^{(j)} x_n^2).$$

For each $N > 0$, define a quadratic polynomial $q_N(x)$ as follows:

$$q_N(x_1,x_2,\ldots,x_n) = p_0 + \frac{1}{2}(x_1,x_2,\ldots,x_n)H_0(x_1,x_2,\ldots,x_n)^T + \sum_{j=3}^{d} \frac{1}{jN^{j-2}} \left(k_1^{(j)}x_1^2 + k_2^{(j)}x_2^2 + \cdots + k_n^{(j)}x_n^2\right). \tag{9}$$

It is clear that the requirements (i) and (ii) of Theorem 2 are satisfied, and the requirement (iv), i.e., the inequality

$$p(x) \le p_0 + \frac{1}{2}x \cdot H_0 \cdot x^T + \sum_{j=3}^{d} |H_j(x)| \le q_N(x)$$

is also true for any $x_1, x_2, \ldots, x_n \in (-1/N, 1/N)$ according to Lemma 1. To see that the requirement (iii) is satisfied for sufficient large N, observe that the Hessian matrix $H_{q_N}(0)$ of $q_N(x)$ at $x = 0$ can be written as $H_0 + 2G_N$, where G_N is the diagonal matrix

$$\begin{pmatrix} g_1(1/N) & & & \\ & g_2(1/N) & & \\ & & \ddots & \\ & & & g_n(1/N) \end{pmatrix},$$

where

$$g_i(y) = \sum_{j=3}^{d} \frac{k_i^{(j)}}{j} \cdot y^{j-2}, \quad i = 1, 2, \cdots, n.$$

Notice that $\lambda_i < 0$, $k_i^{(j)} > 0$ for all i, j $(i = 1, 2, \ldots, n; j = 3, \cdots, d)$, so for each i, the equation

$$\frac{1}{2}\lambda_i + g_i(y) = \frac{1}{2}\lambda_i + \frac{k_i^{(3)}}{3} \cdot y + \cdots + \frac{k_i^{(d)}}{d} \cdot y^{d-2} = 0$$

has a unique positive real root y_i^*. Thus, if the number N satisfies

$$\frac{1}{N} < \min\{y_1^*, y_2^*, \ldots, y_n^*\},$$

then the eigenvalues of $H_0 + 2G_N$, i.e., $\lambda_i + 2g_i(1/N)$ $(i = 1, 2, \ldots, n)$ are all negative, and, therefore, the Hessian matrix of quadratic polynomials $q_N(x)$ is negative-definite, as claimed in (iii).

Theorem 2 is proved. □

Remark 2. Let $\lambda_0 = \max\{\lambda_1, \lambda_2, \ldots, \lambda_n\} < 0$ be the largest eigenvalue of H_0, $C_j = \max\{k_1^{(j)}, k_2^{(j)}, \ldots, k_n^{(j)}\}$ for $j = 3, \cdots, d$, and $1/N$ the smallest positive real root of the following equation:

$$\frac{1}{2}\lambda_0 + \frac{C_3}{3} \cdot \left(\frac{1}{N}\right) + \cdots + \frac{C_d}{d} \cdot \left(\frac{1}{N}\right)^{d-2} = 0.$$

Then, for any $x_2, x_2, \ldots, x_n \in (-1/N, 1/N)$, we have

$$p(x) \leq q_N(x) \leq p(0).$$

3 Local Critical Analysis of Rational and Radical Functions

In this section, we explain how to extend the local critical analysis method to rational functions and certain radical functions of several variables.

3.1 Rational Functions

The method we have described in Theorem 2 can be easily generalized to functions $f(x) = p(x)/q(x)$ where $p(x)$ and $q(x)$ are polynomials of $x \in \mathbb{R}^n$. Let x_0 be a local maximal (or minimal) point such that the Hessian matrix of f at the point x_0 is negative-definite (or positive-definite, respectively). Without loss of generality, we may assume that x_0 is a local minimal point of $f(x)$ and $q(x_0) > 0$. Clearly, if $q(x)$ is positive-definite, then the task of finding a neighborhood $U_0 \subset \mathbb{R}^n$ of x_0 where

$$\frac{p(x)}{q(x)} \geq f(x_0),$$

for all $x \in U_0$ can be simply transformed to finding the neighborhood U_0 where

$$p(x) - f(x_0) \cdot q(x) \geq 0,$$

for $x \in U_0$, which is same as we have done in the previous section. If $q(x)$ is neither positive-definite nor negative-definite in certain known region, we need first to construct such a neighborhood V_0 of x_0 so that $q(x_0) \cdot q(x) > 0$ for all points $x \in V_0$. To implement this work, we have the following theorem.

Theorem 3. *Let $q(x)$ be a polynomial in n variables of degree s, x_0 a point in $\mathbb{R}^n$ with $x_0 = (x_1^*, x_2^*, \ldots, x_n^*)$, and $q(x_0) > 0$,*

$$K_1 = \max\{|\frac{\partial q}{\partial x_i}(x_0)|,\ i = 1, 2, \ldots, n\} > 0,$$
$$K_2 = \max\{|\frac{\partial^2 q}{\partial x_i \partial x_j}(x_0)|, 1 \leq i, j \leq n\} > 0,$$

and

$$K_j = \max\{|\frac{\partial^j q}{\partial x_{i_1} \ldots \partial x_{i_j}}(x_0)|, 1 \leq i_1, \ldots, i_j \leq n\} > 0,$$

for $j = 3, \ldots, s$. Let δ_0 be the unique solution of the equation

$$q(x_0) = K_1 u + \frac{1}{2!} K_2 u^2 + \frac{1}{3!} K_3 u^3 + \cdots + \frac{1}{s!} K_s u^s. \tag{10}$$

Then, the inequality $q(x) > 0$ is valid for any $x = (x_1, x_2, \ldots, x_n)$ with

$$||x - x_0||_1 = |x_1 - x_1^*| + |x_2 - x_2^*| + \cdots + |x_n - x_n^*| < \delta_0.$$

Proof. Let

$$u_1 = x_1 - x_1^*, u_2 = x_2 - x_2^*, \ldots, u_n = x_n - x_n^*,$$

and h_j $(j = 3, \ldots, s)$ the homogeneous polynomials defined by

$$h_j(u_1,\ldots,u_n) = \frac{1}{j!}\left[\sum_{i=1}^{n} u_i \frac{\partial}{\partial x_i'}\right]^j q(x_1',\ldots,x_n')|_{x_1'=x_1^*,\ldots,x_n'=x_n^*}$$
$$= \frac{1}{j!}\sum_{d_1+\cdots+d_n=j}\binom{j}{d_1, d_2, \ldots, d_n}\prod_{i=1}^{n}\left(u_i\frac{\partial}{\partial x_i'}\right)^{d_i} q(x')|_{x'=x_0}.$$

Then, we may expand $q(x)$ in a Taylor series at the point x_0 as follows:

$$q(x_1, x_2, \ldots, x_n) = q(x_1^*, x_2^*, \ldots, x_n^*)$$
$$+\left[u_1\frac{\partial q}{\partial x_1} + u_2\frac{\partial q}{\partial x_2} + \cdots + u_n\frac{\partial q}{\partial x_n}\right]_{x_1=x_1^*,\ldots,x_n=x_n^*}$$
$$+\frac{1}{2!}\left[u_1^2\frac{\partial^2 q}{\partial x_1^2} + 2u_1u_2\frac{\partial^2 q}{\partial x_1\partial x_2} + \cdots + u_n^2\frac{\partial^2 q}{\partial x_n^2}\right]_{x_1=x_1^*,\ldots,x_n=x_n^*}$$
$$+ h_3(u_1, u_2, \ldots, u_n) + \cdots + h_s(u_1, u_2, \ldots, u_n).$$

It is obvious that

$$\texttt{abs}\left(u_1\frac{\partial q}{\partial x_1} + u_2\frac{\partial q}{\partial x_2} + \cdots + u_n\frac{\partial q}{\partial x_n}\right)_{x_1=x_1^*,\ldots,x_n=x_n^*}$$
$$\le K_1(|u_1| + |u_2| + \cdots + |u_n|), \tag{11}$$

$$\texttt{abs}\left[u_1^2\frac{\partial^2 q}{\partial x_1^2} + 2u_1u_2\frac{\partial^2 q}{\partial x_1\partial x_2} + \cdots + u_n^2\frac{\partial^2 q}{\partial x_n^2}\right]_{x_1=x_1^*,\ldots,x_n=x_n^*}$$
$$\le K_2(|u_1|^2 + 2|u_1||u_2| + \cdots + |u_n|^2)$$
$$= K_2(|u_1| + |u_2| + \cdots + |u_n|)^2. \tag{12}$$

For $h_j(u_1, u_2, \ldots, u_n)$ $(j = 3, \ldots, s)$, we have

$$\texttt{abs}\,(h_1(u_1, u_2, \ldots, u_n))$$
$$\le \frac{1}{j!}\sum_{d_1+\cdots+d_n=j}\binom{j}{d_1, d_2, \ldots, d_n}\left(K_j \cdot \prod_{i=1}^{n}(|u_i|)^{d_i}\right)$$
$$= \frac{1}{j!}\cdot K_j \cdot [|u_1| + |u_2| + \cdots + |u_n|]^j\,. \tag{13}$$

Therefore,

$$q(x) \ge q(x_0) - K_1||x - x_0||_1 - \frac{1}{2}K_2||x - x_0||_1$$
$$- \frac{1}{3!}K_3||x - x_0||_1^2 \cdots - \frac{1}{s!}K_s||x - x_0||^s. \tag{14}$$

which immediately implies that $q(x) > 0$ if $||x - x_0||_1 < \delta_0$ and δ_0 is the (unique) real root of the equation (10).

Theorem 3 is proved. □

3.2 Sum of Radicals

Now we consider the radical functions of the following form:

$$f(x) = c_1\sqrt{1 + \frac{p_1(x)}{q_1(x)}} + c_2\sqrt{1 + \frac{p_2(x)}{q_2(x)}} + \cdots + c_k\sqrt{1 + \frac{p_k(x)}{q_k(x)}}, \qquad (15)$$

where $p_j(x)$ and $q_j(x)$ are the polynomials in n variables. We can prove the following result.

Theorem 4. *Assume that $f(x)$ is function defined in (15), $x_0 = 0$, and*

$$p_j(x_0) = 0, \quad q_j(x_0) > 0$$

for $j = 1, 2, \ldots, k$. Then using symbolic computation we can construct a neighborhood U_0 of x_0 and rational functions

$$h(x) = \sum_{j=1}^{k} c_j + \frac{P_1(x)}{Q_1(x)}, \quad g(x) = \sum_{j=1}^{k} c_j + \frac{P(x)}{Q(x)}, \qquad (16)$$

where $P_1(x), Q_1(x), P(x), Q(x)$ are polynomials such that

$$P_1(0) = 0, \quad P(0) = 0,$$

and

$$Q_1(x) > 0, \quad Q(x) > 0, \quad h(x) \le f(x) \le g(x), \qquad (17)$$

for all $x \in U_0$.

To prove this theorem, we need the following Lemma 2 and Lemma 3.

Lemma 2. *For any real number x with $-0.3777 < x < 0.7145$, the following inequality is true:*

$$1 + \frac{1}{2}x - \frac{5}{32}x^2 \le \sqrt{1+x} \le 1 + \frac{1}{2}x - \frac{3}{32}x^2. \qquad (18)$$

□

Lemma 3. *Assume that $p(x)$ and $q(x)$ are the polynomials in n variables $x_1, x_2, \ldots, x_n$, $x_0 = (0, 0, \ldots, 0) \in \mathbb{R}^n$, and*

$$p(x_0) = 0, \quad q(x_0) > 0.$$

Then for any $\varepsilon > 0$, we can find a constant $\delta = \delta(\varepsilon) > 0$ such that

$$\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} < \delta \implies -\varepsilon < \frac{p(x_1, x_2, \ldots, x_n)}{q(x_1, x_2, \ldots, x_n)} < \varepsilon$$

by symbolic computation.

Proof. Indeed, the existence of the $\delta(\varepsilon)$ for each $\varepsilon > 0$ is guaranteed by the continuity of $p(x)/q(x)$ at the point $x_0 = 0$. Here we show that $\delta(\varepsilon)$ can be obtained by symbolic computation. For this purpose, we may assume that in Theorem 3 we have a neighborhood

$$U_0 := \{(x_1, x_2, \ldots, x_n), |x_1| + |x_2| + \cdots + |x_n| < \delta_0\}$$

that satisfies $q(x) > q(0)/2 > 0$ for all $x \in U_0$. Assume $\deg(p) = r$ and

$$p(x) = x \cdot \nabla_{x=0} p(x) + \frac{1}{2} x \cdot H_0 x^T + h_3(x) + \cdots + h_r(x),$$

here h_j are homogeneous polynomials in $x_1, x_2, \ldots, x_n$ for $j = 3, \ldots, r$. Then applying the method described in Lemma 1 and Remark 1 given in the previous section, we can obtain constants $C_j > 0\,(j = 3, \ldots, s)$ so that

$$|h_3(x) + \cdots + h_r(x)| \le \sum_{j=3}^{r} \frac{C_j}{jN^{j-2}} ({x_1}^2 + {x_2}^2 + \cdots + {x_n}^2) \tag{19}$$

for all $x \in \mathbb{R}^n$ with $x_1, x_2, \ldots, x_n \in (-1/N, 1/N)$ for any $N > 0$. Thus, for $(x_1, x_2, \ldots, x_n) \in U_0$, we have $x_1, x_2, \ldots, x_n \in (-\delta_0, \delta_0)$, and inequality (19) implies that

$$|h_3(x) + \cdots + h_r(x)| \le C \cdot ({x_1}^2 + {x_2}^2 + \cdots + {x_n}^2). \tag{20}$$

Here

$$C = \sum_{j=3}^{r} \frac{C_j}{j} \delta_0^{j-2}.$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of H_0, P the orthogonal matrix, (i.e., $P^T P = I$) satisfying

$$H_0 = P^T \cdot \Lambda \cdot P = P^T \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} P,$$

and

$$(x'_1, x'_2, \ldots, x'_n) = (x_1, x_2, \ldots, x_n) P^T.$$

Then we have

$$\begin{aligned} &|x \cdot H_0 \cdot x^T| = |x P^T \cdot \Lambda \cdot P x^T| = |x' \Lambda (x')^T| \\ &= |\lambda_1| {x'}_1^2 + |\lambda_2| {x'}_2^2 + \cdots + |\lambda_n| {x'}_n^2 \\ &\le \sqrt{\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_n^2} \left({x'_1}^2 + {x'_2}^2 + \cdots + {x'_n}^2 \right), \end{aligned}$$

Note that

$$\lambda_1^2 + \lambda_1^2 + \cdots + \lambda_n^2 = \mathrm{tr}(H_0 H_0^T) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial^2 p}{\partial x_i \partial x_j} \frac{\partial^2 p}{\partial x_j \partial x_i} \right)$$
$$= \sum_{i,j=1}^{n} \left(\frac{\partial^2 p}{\partial x_i \partial x_j} \right)^2 = ||H_0||_F^2,$$

and

$${x'_1}^2 + {x'_2}^2 + \cdots + {x'_n}^2 = x' \, {x'}^T = x P^T P x^T = x_1^2 + x_2^2 + \cdots + x_n^2,$$

hence, we get

$$\frac{1}{2}|x \cdot H_0 \cdot x^T| \leq \frac{1}{2}||H_0||_F \cdot \left(x_1^2 + x_2^2 + \cdots + x_n^2 \right). \tag{21}$$

In view of the Cauchy–Schwarz inequality, we have

$$|x \cdot \nabla_{x=0} p(x)| = |x_1 \frac{\partial p}{\partial x_1}(0) + x_2 \frac{\partial p}{\partial x_2}(0) + \cdots + x_n \frac{\partial p}{\partial x_n}(0)|$$
$$\leq ||\nabla_0 p||_2 \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{22}$$

Here

$$||\nabla_0 p||_2 = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial p}{\partial x_i}(0) \right)^2}.$$

Combining (20), (21), and (22), we obtain the following inequality

$$|p(x)| \leq ||\nabla_0 p||_2 \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$
$$+ \left(\frac{1}{2}||H_0||_F + C \right) \left(x_1^2 + x_2^2 + \cdots + x_n^2 \right). \tag{23}$$

Therefore, if we take $\delta(\varepsilon) < \min\{\delta_0/\sqrt{n}, \, \delta_1\}$, where δ_1 is the unique real root of

$$\frac{1}{2}q(0)\varepsilon = ||\nabla_0 p||_2 \, u + \left(\frac{1}{2}||H_0||_F + C \right) u^2,$$

then, from

$$\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} < \delta(\varepsilon),$$

we have

$$|p(x)| < \frac{1}{2}q(0)\varepsilon$$

and $|x_1| + |x_2| + \cdots + |x_n| < \sqrt{n}\delta(\varepsilon) \leq \delta_0$, which implies that $q(x) > q(0)/2 > 0$, and, therefore,

$$|\frac{p(x)}{q(x)}| < \frac{|p(x)|}{\frac{1}{2}q(0)} < \varepsilon,$$

as claimed by Lemma 3. □

Proof (*Proof of Theorem* 4). Without loss of generality, we may assume that $c_j > 0$ for $j = 1, \ldots, l$ and $c_j < 0$ for $j = l+1, \ldots, k$. Then, applying Lemma 3 we can construct a neighborhood U_j so that

$$|\frac{p_j(x)}{q_j(x)}| < 0.3777, \quad q_j(x) > 0$$

for each j $(1 \leq j \leq k)$. Therefore, for point $x \in U_0 := U_1 \cap U_2 \cap \cdots \cap U_k$, we have

$$\begin{aligned} f(x) \leq \sum_{1\leq j\leq l} c_j \left(1 + \frac{p_j(x)}{2q_j(x)} - \frac{3p_j(x)^2}{8q_j(x)^2}\right) \\ + \sum_{l+1\leq j\leq k} c_j \left(1 + \frac{p_j(x)}{2q_j(x)} - \frac{5p_j(x)^2}{8q_j(x)^2}\right) =: g(x). \end{aligned}$$

Let

$$\begin{aligned} P(x) &:= (g(x) - c_1 - c_2 - \cdots - c_k) \cdot Q(x), \\ Q(x) &:= (\texttt{lcm}(q_1(x)q_2(x)\cdots q_k(x)))^2. \end{aligned}$$

Then $Q(x) > 0$ for $x \in U_0$ obviously, $f(x) \leq g(x)$ for $x \in U_0$ as defined, and

$$\begin{aligned} \mathrm{g}(0) = &\sum_{1\leq j\leq l} c_j \left(1 + \frac{p_j(0)}{2q_j(0)} - \frac{3p_j(0)^2}{8q_j(0)^2}\right) \\ &+ \sum_{l+1\leq j\leq k} c_j \left(1 + \frac{p_j(0)}{2q_j(0)} - \frac{5p_j(0)^2}{8q_j(0)^2}\right) \\ = &\, c_1 + c_2 + \cdots + c_k, \end{aligned}$$

therefore, $P(0) = 0$.

The rational function $h(x)$ and the polynomials $P_1(x)$ and $Q_1(x)$ can be constructed by a similar computation. Theorem 4 is proved. □

Our goal is to process the situation when $x_0 = 0$ is a local maximal or minimal point of f. Namely, we wish that the upper-bound rational function $g(x)$ (and the lower-bound rational function $h(x)$, resp.) constructed by Theorem 4 has also taken the point x_0 as the local maximal (minimal, resp.) point if it is a local maximal (minimal, resp.) point of the original radical function $f(x)$, which means, $g(x)$ satisfies the following properties:

– $g'(0) = 0$, and, at best,
– the Hessian matrix $H_g(0)$ is negative-definite,

if x_0 is, for example, a maximal point of $f(x)$. To see this, we have

$$g'(0) = \sum_{j=1}^{k} \left(\frac{1}{2}c_j - \frac{c_j'}{8} \cdot \frac{2p_j(0)}{q_j(0)}\right) \left(\frac{q_j(0)p_j'(0) - q_j'(0)p_j(0)}{q_j(0)^2}\right) = \sum_{j=1}^{k} c_j \cdot \frac{p_j'(0)}{2q_j(0)},$$

here $c_j' = 3c_j$ for $1 \le j \le l$ and $c_j' = 5c_j$ for $l+1 \le j \le k$. Meanwhile, we have

$$f'(0) = \sum_{j=1}^{k} c_j \frac{\left[q_j(0)p_j'(0) - q_j'(0)p_j(0)\right] / \left[q_j(0)^2\right]}{2\left[1 + (p(0)/q(0))^2\right]} = \sum_{j=1}^{k} c_j \cdot \frac{p_j'(0)}{2q_j(0)},$$

which means that if $x_0 = 0$ is a local optimal point of the radical function defined by (15), then it is also a critical point of the upper-bound (or lower bound) rational function $g(x)$ (or $h(x)$, resp.) obtained by Theorem 4.

Remark 3. Notice that $x_0 = 0$ might not be a local maximal point of the upper-bound rational function $g(x)$ even if it is a local maximal point of $f(x)$. To ensure that

$$H_0(f) \text{ is negative-definite} \Longrightarrow H_0(g) \text{ is negative-definite},$$

we may need to refine inequalities (18) of Lemma 2. For example, we may use the following inequality

$$\sqrt{1+x} \le 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 \ (-1 < x < +\infty), \tag{24}$$

for c_j $(1 \le j \le l)$, and the inequality

$$\sqrt{1+x} \ge 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{1}{16}x^4 \ (-0.5161 < x < 3), \tag{25}$$

for c_j $(l+1 \le j \le k)$. The upper-bound rational function $g(x)$ generated by (24) and (25) satisfies $H_0(g) = H_0(f)$ since

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 g}{\partial x_i \partial x_j} \ (1 \le i, j \le n).$$

We omit its proof here.

4 Local Critical Analysis of the Spherical Six-Point Problem

In this section, we discuss the optimization spherical point problem we have mentioned in Sect. 1. Recall that the numerical result says that best arrangement is

$$\Gamma_6 := \{(0,0,1),\ (0,-1,0),\ (1,0,0),\ (0,1,0),\ (-1,0,0),\ (0,0,-1)\}, \tag{26}$$

up to certain rotation of the sphere. We will prove the following theorem.

Theorem 5. *Assume that the six points $P_1, P_2, \ldots, P_6$ are placed on the unit sphere S^2 as follows:*

$$P_1 = (0,0,1), \quad P_2 = (0, -\sqrt{1-z_2^2}, z_2),$$
$$P_3 = (\sqrt{1-y_3^2-z_3^2}, y_3, z_3), \quad P_4 = (x_4, \sqrt{1-x_4^2-z_4^2}, z_4),$$
$$P_5 = (-\sqrt{1-y_5^2-z_5^2}, y_5, z_5), \quad P_6 = (x_6, y_6, -\sqrt{1-x_6^2-y_6^2}),$$

so that

$$-\frac{1}{22.9} \le z_2, y_3, z_3, x_4, z_4, y_5, z_5, x_6, y_6 \le \frac{1}{22.9}, \quad z_2 \ge 0, \tag{27}$$

then

$$\sum_{1\le i<j\le 6} d(P_i, P_j) \le 6 + 12\sqrt{2},$$

and the equality holds if and only if $P_1, P_2, \ldots, P_6$ are congruent to Γ_6.

Proof. Without loss of generality, we may assume that

$$z_2 = \frac{2p}{1+p^2}, \quad y_3 = \frac{2q}{1+q^2+r^2}, \quad z_3 = \frac{2r}{1+q^2+r^2},$$
$$x_4 = \frac{2u}{1+u^2+v^2}, \quad z_4 = \frac{2v}{1+u^2+v^2},$$
$$y_5 = \frac{2s}{1+s^2+t^2}, \quad z_5 = \frac{2t}{1+s^2+t^2},$$
$$x_6 = \frac{2x}{1+x^2+y^2}, \quad y_6 = \frac{2y}{1+x^2+y^2},$$

where

$$-1/45.7 \le p, q, r, s, t, u, v, x, y \le 1/45.7, \quad p \ge 0.$$

Then, we have

$$d(P_i, P_j) = \begin{cases} 2 \cdot \sqrt{1+w_{ij}}, & (i,j) \in \{(1,6),(2,4),(3,5)\}, \\ \sqrt{2}\cdot\sqrt{1+w_{ij}}, & \text{otherwise}, \end{cases}$$

here w_{ij} are rational functions of $p, q, r, s, t, u, v, x, y$, for example,

$$w_{23} = \frac{-2(p^2q + 2\,pr - q)}{(p^2+1)\,(q^2+r^2+1)}, \; w_{24} = \frac{-(p^2 + 2\,pv + u^2 + v^2)}{(p^2+1)\,(u^2+v^2+1)}.$$

Applying inequality (24) we have

$$\sum_{1\le i<j\le 6} d(P_i, P_j) \le 6 + 12\sqrt{2} + G(p, q, r, s, t, u, v, x, y),$$

here

$$G = \sum_{1\le i<j\le 6} c_{ij}\left(\frac{1}{2}w_{ij} - \frac{1}{8}w_{ij}^2 + \frac{1}{16}w_{ij}^3\right)$$

is a rational function and

$$c_{ij} = \begin{cases} 2, & \text{for } (i,j) \in \{(1,6),(2,4),(3,5)\}, \\ \sqrt{2}, & \text{otherwise}. \end{cases}$$

Using `Maple` we obtain $G = P/Q$, where

$$Q = 8\,\left(p^2+1\right)^3 \left(q^2+r^2+1\right)^3 \left(u^2+v^2+1\right)^3 \cdot \left(s^2+t^2+1\right)^3 \left(x^2+y^2+1\right)^3,$$

and the $P = \texttt{numer}(G)$ is polynomial of degree 30 with 543, 609 monomials, of which the least degree is 2. Therefore, we can write P as a sum of 29 homogeneous polynomials as follows:

$$P = H_2 + H_3 + \cdots + H_{30},$$

where H_{30} can be factorized into

$$H_{30} = -11\,p^6\left(x^2+y^2\right)^3\left(u^2+v^2\right)^3\left(s^2+t^2\right)^3\left(q^2+r^2\right)^3 \leq 0.$$

The number of monomial in $H_j(2 \leq j \leq 29)$ are:

$$\begin{aligned}&34, 37, 217, 279, 947, 1221, 3165, 3885, 8142, 9559, 17033,\\&18977, 29766, 30993, 43117, 41763, 51880, 46416, 52178,\\&42108, 42910, 30102, 27244, 16388, 13536, 6080, 4544, 832.\end{aligned}$$

Using `Maple` we can check that the quadratic form H_2 is negative-definite. For simplicity, we show this later.

Assume that $p, q, r, \ldots, x, y \in (-1/N, 1/N)$. Then, applying Lemma 1, we can obtain the following inequalities:

$$\begin{aligned}|H_3| \leq J_3 &= 4\sqrt{2}(26\,p^2 + 35\,q^2 + 31\,r^2 + 35\,s^2 + 31\,t^2\\&\qquad + 38\,u^2 + 43\,v^2 + 38\,x^2 + 44\,y^2)/3N,\\|H_4| \leq J_4 &= \frac{466\sqrt{2}+170}{N^2}\cdot S_9,\\|H_5| \leq J_5 &= \frac{2177}{N^3}\cdot S_9,\ \ldots,\ |H_{29}| \leq J_{29} = \frac{44743}{N^{29}}\cdot S_9,\end{aligned}$$

here

$$S_9 := p^2 + q^2 + r^2 + s^2 + t^2 + u^2 + v^2 + x^2 + y^2,$$

so $J_3, J_4, \ldots, J_{29}$ can be considered as quadratic forms with a parameter N. We will show more information of J_k at the end of this section. Let $J_{30} = 0$. Then $H_{30} \leq J_{30}$, and we can check that if $N > 45.6866$, $H_2 + (J_3 + J_4 + \cdots + J_{30})$ is also a negative-definite quadratic form. Therefore,

$$P = \sum_{k=2}^{30} H_k \leq H_2 + \sum_{k=3}^{30} J_k \leq 0,$$

and

$$\sum_{1\leq i<j\leq 6} d(P_i, P_j) \leq 6 + 12\sqrt{2},$$

for $p, q, \ldots, x, y \in (-1/45.7, 1/45.7)$, also for $P_1, P_2, \ldots, P_6$ that satisfy (27). This proves Theorem 5. □

Now we show that H_2 is negative-definite. We can write H_2 as follows.

$$H_2 = 4(1+\sqrt{2})(p,q,r,s,t,u,v,x,y)A(p,q,r,s,t,u,v,x,y)^T,$$

where

$$A = \begin{bmatrix} -1 & 0 & a & 0 & a & 0 & b & 0 & c \\ 0 & -1 & 0 & b & 0 & c & 0 & 0 & a \\ a & 0 & -1 & 0 & b & 0 & a & -c & 0 \\ 0 & b & 0 & -1 & 0 & -c & 0 & 0 & a \\ a & 0 & b & 0 & -1 & 0 & a & c & 0 \\ 0 & c & 0 & -c & 0 & -1 & 0 & a & 0 \\ b & 0 & a & 0 & a & 0 & -1 & 0 & -c \\ 0 & 0 & -c & 0 & c & a & 0 & -1 & 0 \\ c & a & 0 & a & 0 & 0 & -c & 0 & -1 \end{bmatrix},$$

and

$$a = \sqrt{2}-2,\ b = 1-\sqrt{2},\ c = \sqrt{2}/2-1.$$

The characteristic polynomial of A is

$$\begin{aligned} f(\lambda) &= \left(\lambda^3 + 12\,\lambda\sqrt{2} + 3\,\lambda^2 + 27\sqrt{2} - 15\,\lambda - 38\right) \\ &\times \left(-\lambda^2 + 2\sqrt{2}\,\lambda - 5\,\lambda + 3\sqrt{2} - 5\right) \times \left(\lambda^2 + \lambda + 5\sqrt{2} - 7\right) \\ &\times \left(\lambda - 4 + 3\sqrt{2}\right) \times \left(-\lambda - 4 + \sqrt{2}\right) = 0. \end{aligned}$$

Using `Maple` it is easy to see that $f(\lambda) = 0$ has 9 zeros and all of them are are negative numbers. The largest one is

$$-1/2 + 1/2\sqrt{29 - 20\sqrt{2}} \approx -0.07699\cdots < 0.$$

Therefore, H_2 is a negative-definite quadratic form.

To conclude the paper we show more details about H_k and J_k for $k \geq 3$. As for $4 \leq k \leq 29$, the degree-k homogeneous polynomial H_k has more than 200 monomials, here we only show the cubic homogeneous polynomial H_3 and the construction of J_3 and J_4.

The cubic homogeneous polynomial H_3 has 37 monomials, and all coefficients have a common factor $4\sqrt{2}$.

$$\begin{aligned} H_3 = -4\sqrt{2}(&4\,px^2 + py^2 + 4\,p^2q + 4\,p^2s + p^2y - 4\,u^2y) \\ &-q^2u - 4\,q^2x - qu^2 - 4\,qv^2 + 4\,vx^2 + vy^2 + 4\,ty^2 - v^2y - 4\,r^2u - r^2x \\ &+rx^2 + 4\,ry^2 + s^2u + tx^2 + 4\,s^2x - su^2 - 4\,sv^2 + 4\,t^2u + t^2x \\ &-4\,uvx + 4\,uxy + 4\,qrv - 4\,qry + 4\,qxy - 4\,sxy \\ &-4\,tuv - 4\,pqr - 4\,pst + 4\,ruv + 4\,stv - 4\,sty. \end{aligned}$$

We observe that there are two types of monomials in H_4: those monomials of the form $c{\cdot}w_1^2w_2$ in the first three lines, and those monomials of the form $c{\cdot}w_1w_2w_3$ in the last two lines (printed with underwave), where $w_1, w_2, w_3 \in \{p, q, r, s, t, u, v, x, y\}$ and $w_i \neq w_j$ for $i \neq j$. Notice also that

$$-4\sqrt{2}(\underline{4px^2 + \sqrt{2}py^2}) \leq 0,$$

for $p > 0$. Applying the above inequality to the first two monomials (underlined) of H_3 and the following inequalities

$$cw_1^2w_2 \leq \frac{|c|}{3N}(2w_1^2 + w_2^2),\ cw_1w_2w_3 \leq \frac{|c|}{3N}(w_1^2 + w_2^2 + w_3^2)$$

to the remaining 35 monomials of H_2 of corresponding types, we obtain the upper bound quadratic form of H_3.

$$J_3 = \frac{4\sqrt{2}}{3N}(31p^2 + 35q^2 + 31r^2 + \cdots + 46x^2 + 46y^2).$$

For H_4, the monomials can be classified into five types and for each type we have its corresponding upper bound form as follows:

(1) monomials in the form $c \cdot w_i^4$, which upper bounds are $c'w_i^2/N^2$, with $c' = \max\{0, c\}$;
(2) monomials in the form $c \cdot w_i^2w_j^2$, which upper bounds are $c'(w_i^2 + w_j)/(2N^2)$, with $c' = \max\{0, c\}$;
(3) monomials in the form $c{\cdot}w_i^3w_j$, the corresponding upper bounds are $|c|(3w_i^2 + w_j^2)/(4N^2)$;
(4) monomials in the form $c \cdot w_i^2w_jw_k$, their upper bounds are $|c|(2w_i^2 + w_j^2 + w_k^2)/(4N^2)$;
(5) monomials in the form $c \cdot w_iw_jw_kw_l$,their upper bounds are $|c|(w_i^2 + w_j^2 + w_k^2 + w_l^2)/(4N^2)$.

where $c \in R$ and $w_i, w_j, w_k, w_l \in \{p, q, r, \ldots, x, y\}$. For obtaining tighter upper bound, we have taken

$$c' = \begin{cases} 0, \text{ if } c < 0, \\ c, \text{ otherwise}, \end{cases}$$

in the first two cases. Therefore, we obtain the following result:

$$J_4 = \frac{452\sqrt{2} + 170}{N^2}p^2 + \frac{374\sqrt{2} + 168}{N^2}q^2 + \cdots + \frac{490\sqrt{2} + 72}{N^2}y^2.$$

The largest coefficient of J_4 is $(466\sqrt{2} + 170)/N^2$, thus we have

$$J_4 \leq \frac{466\sqrt{2} + 170}{N^2}(p^2 + q^2 + r^2 + s^2 + t^2 + u^2 + v^2 + x^2 + y^2).$$

Similarly, we have

$$J_k \leq \frac{c_k}{N^{k-2}}(p^2 + q^2 + r^2 + s^2 + t^2 + u^2 + v^2 + x^2 + y^2)$$

for $k = 5, 6, \ldots, 29$, where we can take integer c_k as follows:

$$2177, 9031, 21156, 61636, 121551, 284559, 476083, 938831, 1425542, 2280819,$$
$$3167178, 4135346, 5315958, 5594346, 6708463, 5568363, 6210033, 3953535,$$
$$4035347, 1890122, 1715600, 543585, 421574, 70800, 44743.$$

Clearly,

$$\begin{aligned} P &= H_2 + H_3 + H_4 + \cdots + H_{29} + H_{30} \\ &\leq H_2 + (J_3 + J_4 + \cdots + J_{29}) =: P'(N, p, q, r, \ldots, x, y). \end{aligned}$$

It is easy now to use `Maple` to verify that $P'(N, \cdot)$ is negative-definite.

Acknowledgments. We are grateful to the anonymous reviewers of the ISSAC 2022 and the CASC 2022 for their insightful comments and helpful suggestions, both on mathematics and language aspects, to our manuscripts.

References

1. Berman, J., Hanes, K.: Optimizing the arrangement of points on the unit sphere. Math. Comput. **31**(140), 1006–1008 (1977)
2. Bollobas, B.: An extremal problem for polygons inscribed in a convex curve. Can. J. Math. **19**, 523–528 (1967)
3. Bott, R.: Lectures on morse theory, old and new. Bull. Amer. Math. Soc. (N. S.) **7**, 331–358 (1982)
4. Chattopadhyay, A., Vegter, G., Yap, C.K.: Certified computation of planar Morse-Smale complexes. J. Symb. Comput. **78**, 3–40 (2017)
5. Hou, X.: Spherical Distribution of 5 points with maximal distance sum. Discrete Comput. Geom. **46**, 156–174 (2011)
6. Loi, T.L., Phien, P.: The quantitative Morse theorem. Int. J. Math. Anal. **6**(10), 481–491 (2012)
7. Morse, M.: The Calculus of Variations in the Large, vol. 18. American Mathematical Society Colloquium Publications, New York (1934)
8. Schur, V.: A quantitative version of the Morse lemma and ideal boundary fixing quasiisometries. J. Funct. Anal. **264**(3), 815–836 (2013)
9. Stolarsky, K.B.: Sums of distances between points on a sphere. Proc. AMS **35**(2), 547–549 (1972)
10. Stolarsky, K.B.: Sums of distances between points on a sphere II. Proc. AMS **41**(2), 575–582 (1973)
11. Yomdin, Y.: The geometry of critical and near-critical values of differentiable mappings. Math. Annal. **264**, 495–515 (1983)
12. Zeng, Z., Lu, J., Xu, Y., Wang, Y.: Maximizing the sum of the distances between four points on the unit hemisphere. In: Janičić, P., Kovács, Z. (eds.) Electronic Proceedings in Theoretical Computer Science, vol. 352, pp. 27–40 (2021)
13. Zeng, Z., Zhang, J.: A mechanical proof to a geometric inequality of Zirakzadeh through rectangular partition of polyhedra. J. Syst. Sci. Math. Sci. **30**(11), 1430–1458 (2010)
14. Zirakzadeh, A.: A property of a triangle inscribed in a convex curve. Can. J. Math. **16**, 778–786 (1964)