EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 12, No. 4, 2019, 1360-1370
ISSN 1307-5543 – www.ejpam.com
Published by New York Business Global

Optimal Solution Properties of an Overdetermined System of Linear Equations

Vedran Novoselac[1,*], Zlatko Pavić [1]

[1] *Mechanical Engineering Faculty in Slavonski Brod, University of Osijek, Trg Ivane Brlić Mažuranić 2, 35000 Slavonski Brod, Croatia*

Abstract. The paper considers the solution properties of an overdetermined system of linear equations in a given norm. The problem is observed as a minimization of the corresponding functional of the errors. Presenting the main results of p norm it is shown that the functional is convex. Following the convex properties we examine minimization properties showing that the problem possesses regression, scale, and affine equivariant properties. As an example we illustrated the problem of finding the weighted mean and weighted median of the data.

2010 Mathematics Subject Classifications: 15A06, 65F20, 58K70

Key Words and Phrases: Linear equations, overdetermined systems, equivariance

---

1. Introduction

A system of linear equations denoted as $Ax = b$ where

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n},\ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n,\ b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m. \quad (1)$$

If A is $m \times n$ matrix, with $m > n$, then it is said that the linear system of equations is overdetermined. In general, such a system will have no solution, i.e. it is inconsistent, i.e. $b \notin \mathcal{R}(A)$. Instead the solution with the smallest error $\|b - Ax\|_p$ is observed using some $p \in [1, \infty\rangle$ norme, where the problem is to find a vector $x \in \mathbb{R}^n$ such that

$$\min_{x\in\mathbb{R}^n} \|b - Ax\|_p, \quad p \in [1, \infty\rangle. \quad (2)$$

Figure 1 presents the problem of an overdetermined system of linear equations.

---

[*]Corresponding author.

DOI: https://doi.org/10.29020/nybg.ejpam.v12i4.3517

Email addresses: Vedran.Novoselac@sfsb.hr (V. Novoselac), Zlatko.Pavic@sfsb.hr (Z. Pavić)

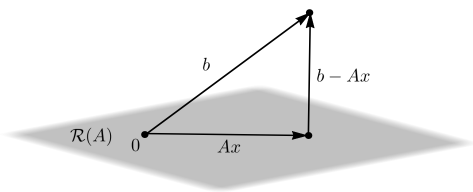


Figure 1: Graphic illustration of an overdetermined system $Ax = b$.

Problem (2) can be observed as a minimization problem of a functional

$$f_p(x) = \|b - Ax\|_p = \left(\sum_{i=1}^{m} |b_i - \sum_{j=1}^{n} a_{ij}x_j|^p \right)^{\frac{1}{p}}, \quad p \in [1, \infty\rangle. \tag{3}$$

Many problems can be presented as an overdetermined system of linear equations such as the problem of determinating linear models for fitting experimental data. It is intended to help researchers fit appropriate curves to their data. Curve fitting, also known as regression analysis, is a common technique for modelling data [13]. The problem of determining n-dimensional hyperplane, in order to have its graph pass as close as possible to given points in sense of p norme, can be presented as an overdetermined system of linear equations [2].

In linear regression analysis it is most frequently assumed that errors, i.e. so called 'outliers', can occur in measured values of the independent variable. In this case, if Euclidian norm ($p = 2$) is used, vector $x \in \mathbb{R}^n$ is obtained in the sense of Least Squares (LS) problem by minimizing (3). In many technical and other applications using the $p = 1$ norm is much more interesting. Because of robust properties of $p = 1$ norm the outliers in data should not affect the obtained results. In literature this problem is known as the Least Absolute Deviation (LAD) problem, and is an efficient method for outlier detection [11, 14].

For that purpose some properties of functional $f_p\colon \mathbb{R}^n \to \mathbb{R}$ are shown in Section 2, especially taking into account the equivariant properties of a solution of an overdetermined system of linear equations. In Section 3, we analyzed equivariant properties of the weighted mean and weighted median of the data, which are a reduced case of an overdetermined system of linear equations, where $p = 1$, $p = 2$ is observed, and $n = 1$ respectively.

2. Properties of the Functional f_p

In this section we present some properties of functional $f_p\colon \mathbb{R}^n \to \mathbb{R}$ in order to prove equivariant properties. Following the Minkowski inequality presented in Theorem 9, which is proven in Section 5, presented in appendix, we give directly the next theorem.

Theorem 1. *Functional $f_p\colon \mathbb{R}^n \to \mathbb{R}$ is convex on $\mathbb{R}^n$.*

The importance of the extremum problems in applied mathematics leads us to the general study of extremum of functional. It is not easy to know the extremum points for differentiable functions, because it is not always possible to solve $\nabla f_p(x) = 0$ to calculate critical points. In the situation when $p = 2$, the problem is known as the LS problem and it always has a solution. In this case x^* minimizes f_2 if and only if x^* solves normal equation system $A^T Ax = A^T b$. It can be shown that if $A \in \mathbb{R}^{m\times n}$ has full rank n, then the LS problem has a unique minimizer. In that case global extremum can be presented as $x^* = A^+ b$, where matrix $A^+ = (A^T A)^{-1} A^T$ is usually called the Moore-Penrose inverse [3]. In most general cases, convex functional f_p has a particularly simple extremal structure, and there are algorithms to calculate extremum points, supposing its existance [15]. Knowing whether or not a local minimum is also global, is one of the most important questions in optimization [7]. The assumption of convexity gives a positive answer to this question as it is stated in the following theorem.

Theorem 2. *A local minimum x^* of a functional $f_p \colon \mathbb{R}^n \to \mathbb{R}$ is always a global extremum.*

Proof. Suppose that x^* is a local minimum of f_p, that is, there is an open neighborhood U of x^* where $f_p(x^*) \le f_p(x)$, $\forall x \in U$. We prove that $f_p(x^*) \le f_p(y^*)$ for arbitrary $y^* \in \mathbb{R}^n$. Consider the convex combination $(1-\lambda)x^* + \lambda y^*$, for $\lambda \in [0,1]$, the convex combination approaches to x^* as $\lambda \to 0$. Therefore for small enough λ, $(1-\lambda)x^* + \lambda y^*$ is in the neighborhood U. Then

$$\begin{aligned} f_p(x^*) &\le f_p((1-\lambda)x^* + \lambda y^*) \\ &\le (1-\lambda) f_p(x^*) + \lambda f_p(y^*). \end{aligned} \tag{4}$$

Rearranging terms, we have $f_p(x^*) \le f_p(y^*)$.

From the Theorem 2, it directly follows that the local minimum of a convex functional is necessarily the global minimum. However, this minimum is not necessarily unique, a sufficient condition is strict convexity. Let us now discuss some equivariant properties of functional f_p. We considered three types of equivariance: regression, scale, and affine equivariance [11]. The next theorem shows that functional f_p possesses these three types of equvivarance.

Theorem 3. *Let $A \in \mathbb{R}^{m\times n}$, $m > n$, $b \in \mathbb{R}^m$, and $x^* \in \mathbb{R}^n$ such that*

$$f_p(x^*) = \min_{x\in\mathbb{R}^n} \|b - Ax\|_p, \quad p \in [1, \infty\rangle. \tag{5}$$

Then

(a) For an arbitrary $v \in \mathbb{R}^n$, vector $x^* + v$ is a solution of

$$\min_{x\in\mathbb{R}^n} \|b + Av - Ax\|_p. \tag{6}$$

(b) For any $c \in \mathbb{R}$, vector $c\,x^*$ is a solution of

$$\min_{x\in\mathbb{R}^n} \|c\,b - Ax\|_p. \tag{7}$$

(c) For a nonsingular matrix $B \in \mathbb{R}^{n\times n}$, vector $B^{-1}x^*$ is a solution of

$$\min_{x\in\mathbb{R}^n} \|b - ABx\|_p. \tag{8}$$

Proof. Let us discuss:

(a) Let y^* be a solution of (6). Then

$$\|b + Av - Ay^*\|_p = \|b - A(y^* - v)\|_p \geq \|b - Ax^*\|_p, \tag{9}$$

whereby the equation is only if $y^* - v = x^*$.

(b) For $c = 0$ the assertion is obvious. Let $c \neq 0$, and y^* such that is a solution of (7). Then

$$\|cb - Ay^*\|_p = |c| \left\| b - A\left(\frac{y^*}{c}\right) \right\|_p \geq |c|\|b - Ax^*\|_p, \tag{10}$$

whereby the equation is only if $\frac{y^*}{c} = x^*$.

(c) Let y^* be a solution of (8). Then

$$\|b - ABy^*\|_p \geq \|b - Ax^*\|_p, \tag{11}$$

whereby the equation is only if $By^* = x^*$.

Corollary 1. *Let $A \in \mathbb{R}^{m\times n}$, $m > n$, $b \in \mathbb{R}^m$, and $x^* \in \mathbb{R}^n$ such that*

$$f_p(x^*) = \min_{x\in\mathbb{R}^n} \|b - Ax\|_p, \quad p \in [1, \infty\rangle. \tag{12}$$

Then for an arbitrary $v \in \mathbb{R}^n$, $c \in \mathbb{R}$, and nonsingular matrix $B \in \mathbb{R}^{n\times n}$, vector $cB^{-1}(x^* + v)$ is a solution of

$$\min_{x\in\mathbb{R}^n} \|c(b + Av) - ABx\|_p. \tag{13}$$

3. Weighted Mean and Weighted Median of the Data

In this section we will illustrate the equivariant properties of the weighted mean and weighted median of the data. As we mentioned, the problem of the weighted mean and weighted median are essentially reduced to solving an overdetermined system of linear equations. Numerous applications of this problem can be found in various branches of applied research, like image processing [10], or methods for outlier detection [11].

Let $a \in \mathbb{R}^m$ be the vector data with corresponding positive vector data weights $w \in \mathbb{R}^m_+$. If we denote by $A = [\sqrt[p]{w_1}, \ldots, \sqrt[p]{w_m}]^T$ and $b = [\sqrt[p]{w_1}a_1, \ldots, \sqrt[p]{w_m}a_m]^T$ in function $f_p \colon \mathbb{R} \to \mathbb{R}$, where $n = 1$, then follows that

$$f_2(x) = \|b - Ax\|_2 = \sqrt{\sum_{i=1}^{m} w_i(a_i - x)^2}, \tag{14}$$

is convex and attains its global minimum on the set $\mathbb{R}$, which is denoted by $x^* = \text{mean}(w, a)$ and called the weighted mean of the data. If $w_1 = \cdots = w_m = 1$, the global minimum of the corresponding functional (14) is denoted by $x^* = \text{mean}(a)$ and called the mean of the data. Analogously, a real number which minimizes function

$$f_1(x) = \|b - Ax\|_1 = \sum_{i=1}^{m} w_i |a_i - x|, \tag{15}$$

is called the weighted median of the data and is denoted by $x^* = \text{med}(w, a)$. If $w_1 = \cdots = w_m = 1$, the global minimum of the corresponding function (15) is denoted by $x^* = \text{med}(a)$ and called the median of the data. In the Figure 2 we present the example of function f_p, for $p = 2$ and $p = 1$ norm. Functions are generated with data vector $a = [1, 2, 3, 4, 5]^T$ and corresponding weights vector $w = [1, 1, 2, 1, 1]^T$ presented in Figure 2(a) for $p = 2$, and for p=1 in Figure 2(b). In Theorem 4 it is shown that the minimum of the function f_2, i.e. weighted mean, always achieved unique minimum. Considering the different data weights $w = [1, 3, 2, 1, 1]^T$, presented in Figure 2(c), it can be seen that the construction of the minimum of f_1, i.e. weighted median, directly depends on it, and achieved its minimum on interval $[a_{(2)}, a_{(3)}]$, in contrast to the situation in Figure 2(b) where the minimum is unique. Also, it can be seen that function f_1 is a piecewise linear function. This property of functional f_1 is considered for finding a minimum, where details are presented in Theorem 5. In the sequel we give solutions for minimizing problems (14) and (15), i.e. for the weighted mean and weighted median of the data.

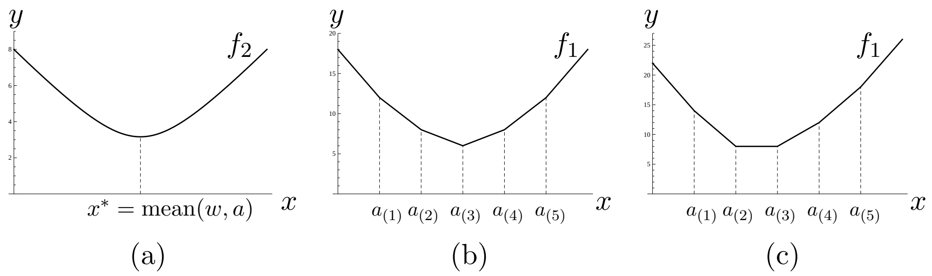


Figure 2: Function $f_p \colon \mathbb{R} \to \mathbb{R}$.

Theorem 4. *Let $a \in \mathbb{R}^m$, $m \geq 2$, be the data vector with corresponding data weights $w \in \mathbb{R}^m_+$. Then*

$$\text{mean}(w, a) = \frac{1}{W} \sum_{i=1}^{m} w_i a_i, \quad W = \sum_{i=1}^{m} w_i. \tag{16}$$

Proof. *Because the functional defined by (14) is derivable, the minimum is attained by finding a solution of* $\frac{\partial f_2(x)}{\partial x} = 0$.

Corollary 2. *Let $a \in \mathbb{R}^m$, $m \geq 2$, be the data vector, then*

$$\operatorname{mean}(a) = \frac{1}{m}\sum_{i=1}^{m} a_i. \tag{17}$$

Theorem 5. *Let $a \in \mathbb{R}^m$, $m \geq 2$, be the data vector with corresponding data weights $w \in \mathbb{R}^m_+$. Let $a_{(1)} \leq a_{(2)} \leq \ldots \leq a_{(m)}$ denote ordered observation and $0 < w_{(1)} \leq w_{(2)} \leq \ldots \leq w_{(m)}$ corresponding weights. Thereby with the denotation*

$$\mathcal{L} := \Big\{ l : \sum_{i=1}^{l} w_{(i)} \leq \frac{W}{2} \Big\}, \quad l \in \{1, \ldots, m\}, \quad W = \sum_{i=1}^{m} w_i, \tag{18}$$

(a) if $\mathcal{L} = \emptyset$, then $\operatorname{med}(w, a) = a_{(1)}$;

(b) if $\mathcal{L} \neq \emptyset$, then with the denotation $l_0 = \max \mathcal{L}$ there holds:

(i) if $\sum_{i=1}^{l_0} w_{(i)} < \frac{W}{2}$, then $\operatorname{med}(w, a) = a_{(l_0+1)}$;

(ii) if $\sum_{i=1}^{l_0} w_{(i)} = \frac{W}{2}$, then $\operatorname{med}(w, a) \in [a_{(l_0)}, a_{(l_0+1)}]$.

Proof. Notice that on each interval

$$\langle -\infty, a_{(1)} \rangle, \ \langle a_{(1)}, a_{(2)} \rangle, \ldots, \langle a_{(m-1)}, a_{(m)} \rangle, \ \langle a_{(m)}, \infty \rangle,$$

function $f_1 \colon \mathbb{R} \to \mathbb{R}$ defined by (15) is linear, and thereby derivable. The slopes of those linear functions are consecutively k_l, $l = 0, \ldots, m$, where

$$k_0 = -\sum_{i=1}^{m} w_i, \quad k_m = \sum_{i=1}^{m} w_i \tag{19}$$

and for $l = 1, \ldots, m-1$

$$k_l = 2\sum_{i=1}^{l} w_{(i)} - \sum_{i=1}^{m} w_i = k_{l-1} + 2w_{(l)}. \tag{20}$$

If $\mathcal{L} = \emptyset$, then for every $l = 1, \ldots, m-1$, is $k_0 < 0 < k_l$. It follows that f_1 is strongly decreasing on $\langle -\infty, a_{(1)} \rangle$ and strongly increasing on $\langle a_{(1)}, \infty \rangle$, therefore the minimum of f_1 is attained for $x^* = a_{(1)}$.

If $\mathcal{L} \neq \emptyset$, then $k_{l+1} - k_l = 2w_{(l+1)} > 0$, and the sequence (k_l) is increasing and

$$k_0 < k_1 < \cdots < k_{l_0} \leq 0 < k_{l_0+1} < \cdots < k_m. \tag{21}$$

If $k_{(l_0)} < 0$, it follows from (21) that f_1 is decreasing on $\langle -\infty, a_{(l_0+1)} \rangle$ and increasing on $\langle a_{(l_0+1)}, \infty \rangle$, therefore the minimum of f_1 is attained for $x^* = a_{(l_0+1)}$

If $k_{(l_0)} = 0$, it follows from (21) that f_1 is decreasing on $\langle -\infty, a_{(l_0)}\rangle$, is constant on $[a_{(l_0)}, a_{(l_0)+1}]$ and increasing on $\langle a_{(l_0+1)}, \infty\rangle$, therefore the minimum of f_1 is attained at every point $x^* \in [a_{(l_0)}, a_{(l_0+1)}]$.

Figure 3(a) present the value of data vector $a = [1,2,3,4,5]^T$. In this situation it is obvious that the weighted mean and weighted median with corresponding data weights $w = [1,1,2,1,1]^T$ of the observed data are equal. Suppose that in some situation two outliers are added, e.g. because of a copying or transmission error. Figure 3(b) displays such a situation, where the two last data have moved up and away from its original position. These are so called outliers, and they have a large influence on the weighted mean, i.e. LS problem, which is quite different from the weighted mean in Figure 3(a).

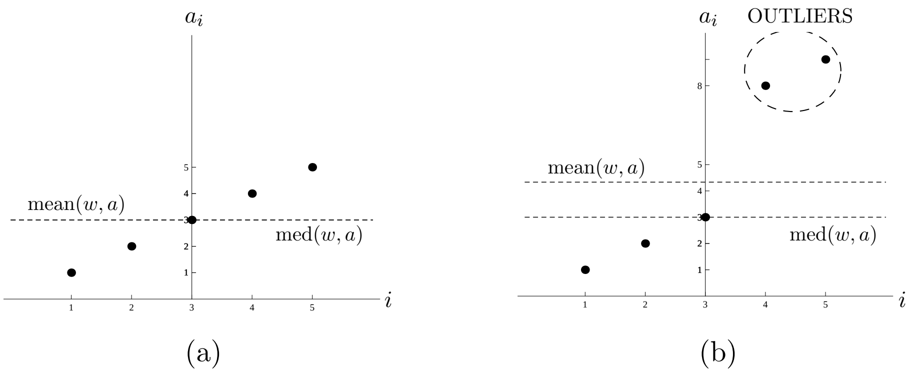


Figure 3: Weighted mean and weighted median: (a) original data; (b) same data as in part (a), but with two outliers.

In the sequel, the corollary is mentioned, which specializes the situation for the case if the weights of the data are not assigned, or if all weights are mutually equal. Also, a description of the pseudo-halving property is mentioned [1], which follows directly from Theorem 5.

Corollary 3. *Let $a \in \mathbb{R}^m$, $m \geq 2$, be the data vector and $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(m)}$ denoted ordered observation. Then follows*

(a) if m is odd ($m = 2l_0 + 1$), $\operatorname{med}(a) = a_{(l_0+1)}$;

(b) if m is even ($m = 2l_0$), $\operatorname{med}(a)$ is every number from the segment $[a_{(l_0)}, a_{(l_0+1)}]$.

Corollary 4. *Let $a \in \mathbb{R}^m$, $m \geq 2$, be the data vector with corresponding data weights $w \in \mathbb{R}^m_+$. Let $a_{(1)} \leq a_{(2)} \leq \ldots \leq a_{(m)}$ denote ordered observation and $0 < w_{(1)} \leq w_{(2)} \leq \ldots \leq w_{(m)}$ corresponding weights. Then there holds that the pseudo-halving property*

$$\sum_{a_{(i)} < x^*} w_{(i)} \leq \frac{W}{2} \quad \textit{and} \quad \sum_{a_{(i)} > x^*} w_{(i)} \leq \frac{W}{2}, \quad W = \sum_{i=1}^{m} w_i. \tag{22}$$

Theorem 6. *Let $a \in \mathbb{R}^m$, $m \geq 2$, be the data vector with corresponding data weights $w \in \mathbb{R}^m_+$. Then for α, β, $\gamma \in \mathbb{R}$, $\alpha > 0$, there holds that*

(a) $\operatorname{mean}(\alpha w, \beta a + \gamma e) = \beta \operatorname{mean}(w, a) + \gamma$;

(b) $\operatorname{med}(\alpha w, \beta a + \gamma e) = \beta \operatorname{med}(w, a) + \gamma$,

where $e = [1, \dots, 1]^T \in \mathbb{R}^m$.

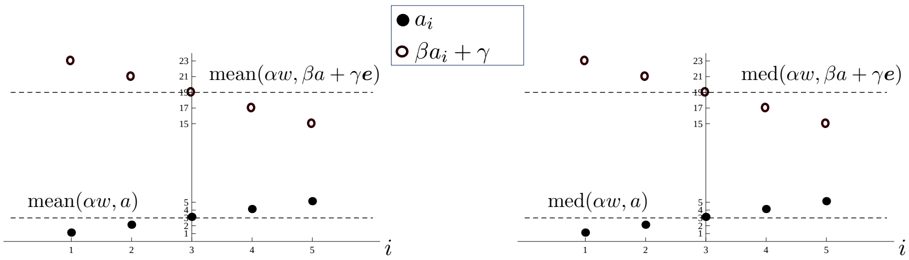
Proof. First we prove (a), while the proof of (b) is going analogue of (a). Notice that equality (a) holds if and only if

$$\operatorname{mean}(\alpha w, a) = \operatorname{mean}(w, a), \quad \text{and} \tag{23}$$

$$\operatorname{mean}(\alpha w, \beta a + \gamma e) = \beta \operatorname{mean}(w, a) + \gamma. \tag{24}$$

Property (23) is trivial to prove. Property (24) follows immediately from Theorem 3, i.e. from regression and scale equivariant property.

The next figure presents the weighted mean and weighted median equivariance properties presented in Theorem 6. For example we observed data vector a and weight vector w from Figure 3(a). Parameter $\alpha > 0$ do not have influence on results of the weighted mean and median. For other parameters we observed case $\beta = -2$, and $\gamma = 25$.



(a) (b)

Figure 4: Equivariance properties: (a) weighted mean; (b) weighted median.

4. Conclusion

Considering the properties of an overdetermined system of linear equations it is established that the solution of the system possesses regression, scale, and affine equivariant properties in observed norm $p \in [1, \infty\rangle$. As an example we observed the problem of finding the weighted mean and weighted median of the data.

5. Appendix - Discrete Forms of Inequalities

A set $S \subseteq \mathbb{R}^n$ is said to be convex if, for all $x, y \in S$ and all $\lambda \in [0,1]$, the point $(1-\lambda)x + \lambda y$ also belongs to S, i.e. $(1-\lambda)x + \lambda y \in S$. The sum $(1-\lambda)x + \lambda y$ is called binomial convex combination. It can be easily seen that if we have m observations $x_1, \ldots, x_m \in S$ in convex set S, and $\lambda_1, \ldots, \lambda_m$ nonnegative number such that $\sum_{i=1}^m \lambda_i = 1$, then $\sum_{i=1}^m \lambda_i x_i \in S$. A point of this type is known as a m-member convex combination of $x_1, \ldots, x_m$.

Let $S \subseteq \mathbb{R}^n$ be convex, a functional $f : S \to \mathbb{R}$ is said to be convex if the inequality

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \tag{25}$$

holds for all points $x, y \in S$ and coefficient $\lambda \in [0,1]$. If the inequality (25) is strict for all $x, y \in S$, then $f(x)$ is called strictly convex.

Using mathematical induction, the inequality in formula (25) can be extended to m-membered convex combinations.

Theorem 7. (Discrete form of Jensen's inequality) *Let $S \subseteq \mathbb{R}^n$ be a convex set, let $x_i \in S$ be points, and let $\lambda_i \in [0,1]$ be coefficients such that $\sum_{i=1}^{m} \lambda_i = 1$. Then each convex function $f : S \to \mathbb{R}$ satisfies the inequality*

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i). \tag{26}$$

Theorem 8. (Discrete form of Hölder's inequality) *Let $x, y \in \mathbb{R}^n$ be points, and let $p, q \in \langle 0, \infty \rangle$ be numbers such that $1/p + 1/q = 1$. Then we have the inequality*

$$\sum_{i=1}^{m} |x_i y_i| \leq \left(\sum_{i=1}^{m} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{m} |y_i|^q\right)^{\frac{1}{q}}. \tag{27}$$

Proof. Assuming that all points y_i are different from zero, formula (27) can be obtained from formula (26) as follows. Using the points

$$|x_i||y_i|^{-\frac{q}{p}}$$

as x_i, the coefficients

$$\lambda_i = \frac{|y_i|^q}{\sum_{i=1}^{m} |y_i|^q},$$

and the convex function x^p, we get

$$\left(\frac{1}{\sum_{i=1}^{m} |y_i|^q} \sum_{i=1}^{m} |y_i|^q |x_i||y_i|^{-\frac{q}{p}}\right)^p \leq \frac{1}{\sum_{i=1}^{m} |y_i|^q} \sum_{i=1}^{m} |y_i|^q \Big(|x_i||y_i|^{-\frac{q}{p}}\Big)^p.$$

Since $q - q/p = 1$, it follows that

$$\left(\frac{1}{\sum\limits_{i=1}^{m}|y_i|^q}\sum_{i=1}^{m}|x_iy_i|\right)^p \leq \frac{1}{\sum\limits_{i=1}^{m}|y_i|^q}\sum_{i=1}^{m}|x_i|^p.$$

Taking the p-th root, multiplying by $\sum_{i=1}^{m}|y_i|^q$, and using the exponent $1/q$ instead of $1-1/p$, we achieve the inequality in formula (27).

If some y_j is equal to zero, then $x_jy_j = 0$ does not increase the left side of formula (27), but $x_j \neq 0$ increases the right side.

Utilizing the vectors $x = [x_1, \dots, x_n]^T$, $y = [y_1, \dots, y_n]^T$ and $z = [x_1y_1, \dots, x_ny_n]^T$, formula (27) can be expressed by the norms,

$$\|z\|_1 \leq \|x\|_p\|y\|_q. \tag{28}$$

Theorem 9. (Discrete form of Minkowski's inequality) *Let $x, y \in \mathbb{R}^n$ be points, and let $p \in [1,\infty)$ be a number. Then we have the inequality*

$$\left(\sum_{i=1}^{m}|x_i+y_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{m}|x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{m}|y_i|^p\right)^{\frac{1}{p}}. \tag{29}$$

Proof. If all x_i and y_i are equal to zero, then formula (29) trivially holds. If $p = 1$, then the inequality in formula (29) follows from the simple triangle inequality $|x_i + y_i| \leq |x_i| + |y_i|$. If some of the points are different from zero, and if $p > 1$, the inequality in formula (29) can be derived by using the simple triangle inequality, and the inequality in formula (27). In this intention, we have

$$\sum_{i=1}^{m}|x_i+y_i|^p \leq \sum_{i=1}^{m}|x_i||x_i+y_i|^{p-1} + \sum_{i=1}^{m}|y_i||x_i+y_i|^{p-1}$$

$$\leq \left(\sum_{i=1}^{m}|x_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{m}|x_i+y_i|^{(p-1)q}\right)^{\frac{1}{q}} + \left(\sum_{i=1}^{m}|y_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{m}|x_i+y_i|^{(p-1)q}\right)^{\frac{1}{q}}$$

$$= \left(\left(\sum_{i=1}^{m}|x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{m}|y_i|^p\right)^{\frac{1}{p}}\right)\left(\sum_{i=1}^{m}|x_i+y_i|^p\right)^{\frac{1}{q}}$$

because $(p-1)q = p$. Dividing by $\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{1/q}$, and putting $1/p$ instead of $1-1/q$, we obtain the inequality in formula (29).

Using p norms of the vectors x, y and $x+y$, the inequality in formula (29) takes the form

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p. \tag{30}$$

References

[1] P. Bloomfield, and W. Steiger. *Least Absolute Deviations: Theory, Applications and Algorithms*. Birkhauser, Boston, 1983.

[2] J. A. Cadzow. Minimum ℓ_1, ℓ_2, and ℓ_∞ norm approximate solutions to an overdetermined system of linear equations. *Digital Signal Processing*, 12(4): 524-560, 2002.

[3] M. Fiedler, J. Nedoma, J. Ramik, J. Rohn, and K. Zimmermann. *Linear optimization problems with inexact data*. Springer, 2006.

[4] J. Mićić, Z. Pavić, and J. Pečarić. The inequalities for quasiarithmetic means. *Abstract and Applied Analysis*, 2012: 1-25, 2012.

[5] T. Needham. A visual explanation of Jensen's inequality. *American Mathematical Monthly*, 100(8): 768-771, 1993.

[6] C. P. Niculescu, and L. E. Persson. *Convex Functions and Their Applications*. Canadian Mathematical Society, Springer, New York, USA, 2006.

[7] M. R. Osborne. *Finite Algorithms in Optimization and Data Analysis*. Departmant of Statistics, Australian National University, Camberra, John Wiley, 1985.

[8] Z. Pavić, J. Pečarić, and I. Perić. Integral, discrete and functional variants of Jensen's inequality. *Journal of Mathematical Inequalities*, 5(2): 253-264, 2011.

[9] J. E. Pečarić, F. Proschan, and Y. L. Tong. *Convex Functions, Partial Orderings, and Statistical Applications*. Academic Press, New York, USA, 1992.

[10] I. Pitas. *Digital Image Processing Algorithms and Applications*. John Wiley & Sons, 2000.

[11] P. J. Rousseeuw, and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 2003.

[12] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, USA, 1987.

[13] V. Sit, and M. Poulin-Costello. *Catalogue of curves for curve fitting*. Biometrics information handbook series, no. 4, 1994.

[14] I. Vazler, K. Sabo, and R. Scitovski. Weighted median of the data in solving least absolute deviations problems. *Communications in Statistics-Theory and Methods*, 41(8): 1455-1465, 2012.

[15] G. A. Watson. *Aproximation Theory and Numerical Methods*. John Wiley & Sons, 1980.