Analysis II for Statisticians - SS16

Iosif Petrakis
petrakis@math.lmu.de

June 24, 2016

These notes include part of the material discussed in our Exercises-session for the lecture course "Analysis II for Statisticians" of Prof. Dr. Tomasz Cieslak.

Please feel free to send me your comments and suggestions regarding these notes.

I. Symmetric matrices

Definition 1. *If* $(X, +, 0, \cdot, 1)$ *is a vector space, a scalar product (or an inner product) on* X *is a map* $<,>: X \times X \to \mathbb{R}$ *satisfying, for every* $x, y, z \in X$ *and* $\lambda \in \mathbb{R}$*, the following properties:*
(i) $< x, x > \geq 0$.
(ii) $< x, x > = 0 \to x = 0$.
(iii) $< x, y > = < y, x >$
(iv) $< x, y + z > = < x, y > + < x, z >$.
(v) $< \lambda x, y > = < x, \lambda y > = \lambda < x, y >$.

It is immediate to check that the map $<,>: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, defined by

$$\begin{aligned} < x, y > &:= x^{\mathrm{T}} y \\ &= \sum_{i=1}^{n} x_i y_i, \end{aligned}$$

is a scalar product on the vector space $\mathbb{R}^n$. Note that in the expression $x^{\mathrm{T}} y$ we consider the elements x, y of $\mathbb{R}^n$ as column vectors i.e., $n \times 1$–matrices, therefore x^{T}, the transpose matrix of x, is an $1 \times n$–matrix. Hence the multiplication $x^{\mathrm{T}} y$ between an $1 \times n$–matrix and an $n \times 1$–matrix is well defined and results to an 1×1–matrix, the real number $\sum_{i=1}^{n} x_i y_i$.

Definition 2. *If* $A = (a_{ij})$ *is an* $n \times m$*–matrix and* $B = (b_{jk})$ *is an* $m \times k$*–matrix, their product* $AB = (c_{ik})$ *is an* $n \times k$*–matrix, where*

$$c_{ik} := \sum_{i=1}^{n} a_{ij} b_{jk}.$$

We also use the notation

$$M^n(\mathbb{R}) := \{A \mid A \text{ is an } n \times n\text{–matrix over } \mathbb{R}\}.$$

Definition 3. *If* $A = (a_{ij})$ *is an* $n \times m$*-matrix, its transpose* $A^{\mathrm{T}} = (b_{ji})$ *is an* $m \times n$*–matrix, where* $b_{ji} = a_{ij}$*. Moreover, if* $A = (a_{ij}) \in M^n(\mathbb{R})$*, its trace* $\mathrm{Tr}(A)$ *is defined by*

$$\mathrm{Tr}(A) := \sum_{i=1}^{n} a_{ii}.$$

One basic property of the transpose of a matrix that we'll use is that

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}.$$

Exercise 4. *Let $S^n(\mathbb{R})$ be defined by*

$$S^n(\mathbb{R}) := \{A \in M^n(\mathbb{R}) \mid A \textit{ is symmetric}\}.$$

(i) *Show that $S^n(\mathbb{R})$ is a vector space.*
(ii) *Show that the map defined by*

$$< A, B >:= \mathrm{Tr}(AB),$$

for every $A, B \in S^n(\mathbb{R})$, is a scalar product on $S^n(\mathbb{R})$.

Blatt 1, Aufgabe 4

If $A \in M^2(\mathbb{R})$ such that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

it is easy to see that

$$|ad - bc|^2 \leq (a^2 + c^2)(b^2 + d^2) \leftrightarrow (ab + cd)^2 \geq 0.$$

The general case for some $n > 2$ is treated as follows. Let

$$A = (\sigma_1 \ldots \sigma_n).$$

By the Gram-Schmidt process there exist $b_1, \ldots, b_n \in \mathbb{R}^n$ such that
(i) $||b_i|| = 1$, for every i.
(ii) $< b_i, b_j >= 0$, for every $i \neq j$.
(iii) $\forall_{x\in\mathbb{R}^n} \exists_{\lambda_1,\ldots,\lambda_n\in\mathbb{R}}(x = \sum_{i=1}^{n} \lambda_i b_i)$.
(iv) $\mathrm{span}\{\sigma_1, \ldots, \sigma_k\} = \mathrm{span}\{b_1, \ldots, b_k\}$, for every $1 \leq k \leq n$.

Consequently, the $n \times n$–matrix

$$B = (b_1 \ldots b_n)$$

is orthogonal and $B^{\mathrm{T}}B = BB^{\mathrm{T}} = \mathbb{I}_n$, where $\mathbb{I}_n$ is the unit element of $M^n(\mathbb{R})$. Moreover, if $x \in \mathbb{R}^n$, we have that

$$x = \sum_{k=1}^{n} < x, b_k > b_k, \qquad (1)$$

since

$$\begin{aligned}
< x, b_k =&< \sum_{i=1}^{n} \lambda_i b_i, b_k > \\
&= \sum_{i=1}^{n} < \lambda_i b_i, b_k > \\
&= \sum_{i=1}^{n} \lambda_i < b_i, b_k > \\
&= \lambda_k < b_k, b_k > \\
&= \lambda_k.
\end{aligned}$$

Using (1) we also get that

$$||x||^2 = \sum_{k=1}^{n} | < x, b_k > |^2, \qquad (2)$$

since

$$\begin{aligned}
||x||^2 &=< x, x > \\
&=< \sum_{k=1}^{n} < x, b_k > b_k, \sum_{k=1}^{n} < x, b_k > b_k > \\
&= \sum_{k=1}^{n} << x, b_k > b_k, < x, b_k > b_k > \\
&= \sum_{k=1}^{n} < x, b_k >^2 < b_k, b_k > \\
&= \sum_{k=1}^{n} < x, b_k >^2 \\
&= \sum_{k=1}^{n} | < x, b_k > |^2.
\end{aligned}$$

By (iv) each σ_k has a shorter expansion than the one found in (1), since

$$\sigma_k \in \text{span}\{b_1, \ldots, b_k\} \leftrightarrow \sigma_k = \sum_{j=1}^{k} \mu_j b_j \leftrightarrow \sigma_k = \sum_{j=1}^{k} < \sigma_k, b_j > b_j. \qquad (3)$$

Next we define the matrix $C = (c_{kl})$ by

$$c_{kl} = \begin{cases} <\sigma_l, b_k> & , \text{ if } 1 \leq k \leq l \\ 0 & , \text{ if } l < k \leq n. \end{cases}$$

Clearly we have that

$$C = \begin{pmatrix} <\sigma_1, b_1> & <\sigma_2, b_1> & \ldots & <\sigma_n, b_1> \\ 0 & <\sigma_2, b_2> & \ldots & <\sigma_n, b_2> \\ 0 & 0 & \ldots & <\sigma_n, b_3> \\ . & . & \ldots & \ldots \\ . & . & \ldots & \ldots \\ 0 & 0 & \ldots & <\sigma_n, b_n> \end{pmatrix}$$

i.e., C is an upper triangular matrix, and because of (3) we get that

$$A = BC.$$

Since B is orthogonal, and since the determinant of a triangular matrix is the product of its diagonal elements, we have that

$$\begin{aligned} \det(A)^2 &= \det(A^{\mathrm{T}}A) \\ &= \det((BC)^{\mathrm{T}}BC) \\ &= \det((C^{\mathrm{T}}B^{\mathrm{T}})BC) \\ &= \det(C^{\mathrm{T}}(B^{\mathrm{T}}B)C) \\ &= \det(C^{\mathrm{T}}\mathbb{I}_nC) \\ &= \det(C^{\mathrm{T}}C) \\ &= \det(C)^2 \\ &= \prod_{k=1}^{n} | <\sigma_k, b_k> |^2 \\ &\leq \prod_{k=1}^{n} \sum_{i=1}^{n} | <\sigma_i, b_k> |^2 \\ &= \prod_{k=1}^{n} ||\sigma_k||^2. \end{aligned}$$

Note that by the previous inequality we get the following necessary and sufficient condition for getting the equality in the main inequality of the exercise. Namely,

$$\det(A)^2 = \prod_{k=1}^{n} ||\sigma_k||^2 \leftrightarrow \forall_{k\in\{1,\ldots,n\}}(| < \sigma_k, b_k > |^2 = \sum_{i=1}^{n} | < \sigma_i, b_k > |^2)$$
$$\leftrightarrow \forall_{k\in\{1,\ldots,n\}}(\sigma_k =< \sigma_k, b_k > b_k)$$
$$\leftrightarrow \forall_{i\neq j}(\sigma_i \perp \sigma_j),$$

since the vectors $b_1, \ldots, b_n$ are pairwise perpendicular to each other.

Question 5. *Can we assert that*

$$|\det(A)| \leq \prod_{1}^{n} ||\tau_i||,$$

where τ_i are the row-vectors of A?

Blatt 2, Aufgabe 4

Next follows the material necessary to a complete presentation of the solution.

Definition 6. *If $I \subseteq \mathbb{R}$ and $f : I \to \mathbb{R}^n$, then there are functions $f_1, \ldots, f_n : I \to \mathbb{R}$ such that we can write f as*

$$f = (f_1, \ldots, f_n),$$

$$f(t) = (f_1(t), \ldots, f_n(t)),$$

for every $t \in I$. We say that f is differentiable at $t \in I$, if $f_1, \ldots, f_n$ are differentiable at t and

$$f'(t) := ({f_1}'(t), \ldots, {f_n}'(t)).$$

We say that f is differentiable on I, if it is differentiable at every $t \in I$.

It is straightforward to show the following proposition.

Proposition 7. *Let $g, h : I \to \mathbb{R}$ be differentiable on I, $P \in \mathbb{R}^n$ and $\delta, \varepsilon : I \to \mathbb{R}^n$ differentiable on I. The following hold:*
(i) *The function $\gamma : I \to \mathbb{R}^n$, defined by*

$$\gamma(t) := g(t)P, \quad t \in I,$$

is differentiable on I and

$$\gamma'(t) := g'(t)P, \quad t \in I.$$

(ii) *$\delta + \varepsilon$ is differentiable on I and*

$$(\delta + \varepsilon)'(t) := \delta'(t) + \varepsilon'(t), \quad t \in I.$$

(iii) *The function $f : I \to \mathbb{R}^n$, defined by*

$$f(t) :=< g(t), h(t) >, \quad t \in I,$$

is differentiable on I and

$$f'(t) :=< g'(t), h(t) > + < g(t), h'(t) >, \quad t \in I.$$

(iv) *If $A \in M^n(\mathbb{R})$, the function $f : I \to \mathbb{R}^n$, defined by*

$$f(t) := A\delta(t), \quad t \in I,$$

is differentiable on I and

$$f'(t) := A\delta'(t), \quad t \in I.$$

Definition 8. *We denote by*

$$S^{n-1} := \{x \in \mathbb{R}^n \mid ||x|| = 1\},$$

the unit sphere of $\mathbb{R}^n$.

For example,

$$S^1 := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

is the standard unit circle. In Exercise 1 of Blatt 2 we showed that S^1 is a compact subset of $\mathbb{R}^2$, and similarly one shows that S^{n-1} is a compact subset of $\mathbb{R}^n$.

Definition 9. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $P \in \mathbb{R}^n$. We say that P is a maximum for f on S^{n-1}, if $P \in S^{n-1}$ i.e., $||P|| = 1$, and if*

$$\forall_{x \in S^{n-1}}(f(x) \le f(P)).$$

Note that P is not necessarily unique, while unique is the maximum value $f(P)$. If f is continuous, then f has always a maximum on S^{n-1}, since S^{n-1} is compact. The next proposition explains why the function considered in Exercise 4 of Blatt 2 is continuous, therefore it is meaningful to talk about its maximum value on S^{n-1}.

If (X, d) is a metric space, $(x_n)_{n=1}^{\infty} \subset X$ and $x \in X$, we use the notation

$$x_n \xrightarrow{n} x := \lim_{n\to\infty} x_n = x.$$

Proposition 10. ***(i)*** *If* $< . >: X \times X \to \mathbb{R}$ *is a scalar product on the vector space* X*, then* $< . >$ *is a continuous functions i.e.,*

$$(x_n, y_n) \xrightarrow{n} (x, y) \ \Rightarrow < x_n, y_n > \xrightarrow{n} < x, y >,$$

for every $(x_n)_{n=1}^{\infty}, (y_n)_{n=1}^{\infty} \subset X$ *and* $x, y \in X$*. Note that*

$$(x_n, y_n) \xrightarrow{n} (x, y) \ \Leftrightarrow ||x_n - x|| \xrightarrow{n} 0 \ \wedge \ ||y_n - y|| \xrightarrow{n} 0.$$

(ii) *If* $A \in M^n(\mathbb{R})$*, then the function* $g : \mathbb{R}^n \to \mathbb{R}^n$ *defined by*

$$g(x) := Ax,$$

for every $x \in \mathbb{R}^n$*, is continuous.*
(iii) *If* $A \in M^n(\mathbb{R})$*, then the function* $f : \mathbb{R}^n \to \mathbb{R}$ *defined by*

$$f(x) :=< x, Ax >,$$

for every $x \in \mathbb{R}^n$*, is continuous.*

Proof. (i) With the use of the Cauchy-Schwarz inequality.
(ii) First you need to unfold the multiplication Ax.
(iii) Use (i) and (ii). □

Theorem 11. *Let* $A \in S^n(\mathbb{R})$ *and* $f : \mathbb{R}^n \to \mathbb{R}$ *defined by*

$$f(x) :=< x, Ax >,$$

for every $x \in \mathbb{R}^n$*. If* P *is a maximum for* f *on* S^{n-1}*, then* P *is an eigenvector of* A*.*

Proof. We consider the set

$$Y = \{y \in \mathbb{R}^n \mid < y, P >= 0\} = \{\lambda P \mid \lambda \in \mathbb{R}\}^{\perp},$$

which is a subspace of $\mathbb{R}^n$ of dimension $n-1$, since

$$\dim(Y) + \dim(Y^{\perp}) = n.$$

Let $y \in Y$ such that $||y|| = 1$, and $\gamma_y : [-1, 1] \to \mathbb{R}^n$ is defined by

$$\gamma_y(t) := (\cos t)P + (\sin t)y, \quad t \in [-1, 1].$$

We show the following:

(i) $\gamma_y(t) \in S^{n-1}$.
(ii) $\gamma_y(0) = P$.
(iii) ${\gamma_y}'(0) = y$.
(iv) γ_y is a curve on S^{n-1} passing through P and the direction of γ_y at 0 is the direction of y.

(i) We have that

$$\begin{aligned} ||\gamma_y(t)||^2 &= < \gamma_y(t), \gamma_y(t) > \\ &= < (\cos t)P + (\sin t)y, (\cos t)P + (\sin t)y > \\ &= (\cos^2 t)||P||^2 + (\sin^2 t)||y||^2 \\ &= 1. \end{aligned}$$

(ii) $\gamma_y(0) = (\cos 0)P + (\sin 0)y = 1P = P$.
(iii) Using Proposition 7 we get ${\gamma_y}'(t) = (-\sin t)P + (\cos t)y$, therefore ${\gamma_y}'(0) = y$.
(iv) This is (i)-(iii) in words.

Next we define the function $g : [-1, 1] \to \mathbb{R}$ by

$$g(t) := f(\gamma_y(t)) = < \gamma_y(t), A\gamma_y(t) >, \quad t \in [-1, 1].$$

By Proposition 7(iii) and (iv), and by the fundamental property of symmetric matrices

$$< x, Ay > = < Ax, y >$$

we have that

$$\begin{aligned} g'(t) &= < {\gamma_y}'(t), A\gamma_y(t) > + < \gamma_y(t), (A\gamma_y(t))' > \\ &= < {\gamma_y}'(t), A\gamma_y(t) > + < \gamma_y(t), A{\gamma_y}'(t) > \\ &= < {\gamma_y}'(t), A\gamma_y(t) > + < A\gamma_y(t), {\gamma_y}'(t) > \\ &= < {\gamma_y}'(t), A\gamma_y(t) > + < {\gamma_y}'(t), A\gamma_y(t) > \\ &= 2 < {\gamma_y}'(t), A\gamma_y(t) > . \end{aligned}$$

Since $f(P)$ is a maximum and $g(0) = f(\gamma_y(0)) = f(P)$, we get

$$\begin{aligned} g'(0) = g'(f(P)) = 0 &\leftrightarrow 2 < {\gamma_y}'(0), A\gamma_y(0) >= 0 \\ &\leftrightarrow 2 < y, AP >= 0 \\ &\rightarrow AP \perp y, \quad y \in Y \\ &\leftrightarrow AP \in Y^{\perp} \\ &\leftrightarrow AP = \lambda P, \end{aligned}$$

for some $\lambda \in \mathbb{R}$. □

The next corollary is exactly Exercise 4 of Blatt 2.

Corollary 12. *Let $A \in S^n(\mathbb{R})$ and $f : \mathbb{R}^n \to \mathbb{R}$ defined by*

$$f(x) :=< x, Ax >,$$

for every $x \in \mathbb{R}^n$. The maximum value of f on S^{n-1} is the maximal eigenvalue of A.

Proof. Let λ be a non-zero eigenvalue of A (it always exists, since A is symmetric). Let P is an eigenvector of A with eigenvalue λ such that $||P|| = 1$ (why can I always tale $P \in S^{n-1}$?). Then we have

$$\begin{aligned} f(P) &=< P, AP > \\ &=< P, \lambda P > \\ &= \lambda < P, P > \\ &= \lambda ||P||^2 \\ &= \lambda. \end{aligned}$$

By the previous theorem the maximum of f on S^{n-1} occurs at an eigenvector, therefore by the previous equality this maximum value is the maximum eigenvalue of A. □

II. Convexity of C^2-functions

(For Exercise 3, Blatt 7)

First we give some necessary definitions.

Definition 13. *Let $A \in M^n(\mathbb{R})$ be a symmetric matrix. A minor Δ_k of order k of A is called principal, if it is obtained by deleting $n-k$ rows and the $n-k$ columns with the same numbers. The leading principal minor D_k of A of order k is the minor of order k obtained by deleting the last $n-k$ rows and columns of A.*

For example, if $n = 2$ and A is given by

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

it has the following principal minors $\Delta_1 = a$ and $\Delta_2 = c$ of order one and the principal minor $\Delta_2 = ac - b^2$ of order two. Moreover, it has the leading principal minor $D_1 = a$ of order one and the leading principal minor $D_2 = ac - b^2$ of order two.

Note that there are

$$\frac{n!}{k!(n-k)!}$$

principal minors of order k.

Theorem 14. *Let $A \in M^n(\mathbb{R})$ be a symmetric matrix.*

(i) *A is positive definite if and only if $D_k > 0$, for all leading principal minors D_k of order k of A, where $1 \leq k \leq n$.*

(ii) *A is negative definite if and only if $(-1)^k D_k > 0$, for all leading principal minors D_k of order k of A, where $1 \leq k \leq n$.*

(iii) *A is positive semi-definite if and only if $\Delta_k \geq 0$, for all principal minors Δ_k of order k of A, where $1 \leq k \leq n$.*

(iv) *A is negative semi-definite if and only if $(-1)^k \Delta_k \geq 0$, for all leading principal minors Δ_k of order k of A, where $1 \leq k \leq n$.*

Definition 15. *If $U \subseteq \mathbb{R}^n$ is open, a function $f : U \to \mathbb{R}$ is called C^2, if the partial derivatives on U*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

exist and are continuous.

Theorem 16. *If $U \subseteq \mathbb{R}^n$ is open and $f : U \to \mathbb{R}$ is C^2, then its Hessian matrix $H_f(x)$, $x \in U$, is symmetric.*

Theorem 17. *Let $U \subseteq \mathbb{R}^n$ be open and convex, and let $f : U \to \mathbb{R}$ be C^2.*

(i) *f is convex in U if and only if H_f is positive semi-definite in U.*

(ii) *f is concave in U if and only if H_f is negative semi-definite in U.*

(iii) *If H_f is positive definite in U, then f is strictly convex in U.*

(iv) *If H_f is negative definite in U, then f is strictly concave in U.*

For the part (ii) of Exercise 3 one needs to use the following simple fact.

Proposition 18. *Let $U \subseteq \mathbb{R}^n$ be convex and $f : U \to \mathbb{R}$ be continuous. If f is (strictly) convex in U, then f is convex in the closure $\overline{U}$ of U.*

Aufgabe 4, Blatt 7

Seien $p, q \in (1, +\infty)$, so dass

$$\frac{1}{p} + \frac{1}{q} = 1,$$

und seien $x_1, \dots, x_k, y_1, \dots, y_k \in \mathbb{R}$. Zeigen Sie, dass die folgenden Gleichungen stimmen.

(i) $$\sum_{i=1}^{k} (|x_i| + |y_i|)^p = \sum_{i=1}^{k} (|x_i| + |y_i|)^{p-1} |x_i| + \sum_{i=1}^{k} (|x_i| + |y_i|)^{p-1} |y_i|.$$

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

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

Proof. (i) First not that if $a, b \in \mathbb{R}$, then

$$\begin{aligned}(|a| + |b|)^p &= (|a| + |b|)^{p-1} (|a| + |b|) \\ &= (|a| + |b|)^{p-1} |a| + (|a| + |b|)^{p-1} |b|.\end{aligned}$$

Hence

$$\sum_{i=1}^{k}(|x_i|+|y_i|)^p = \sum_{i=1}^{k}[(|x_i|+|y_i|)^{p-1}|x_i| + (|x_i|+|y_i|)^{p-1}|y_i|]$$
$$= \sum_{i=1}^{k}(|x_i|+|y_i|)^{p-1}|x_i| + \sum_{i=1}^{k}(|x_i|+|y_i|)^{p-1}|y_i|.$$

(ii) The Hölder inequality is the following:

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

Since $(p-1)q = p$, by Hölder inequality we get

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

(iii) Because of (ii) we also have

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

Hence by (i) we get

$$\sum_{i=1}^{k}(|x_i|+|y_i|)^p \leq \left(\sum_{i=1}^{k}(|x_i|+|y_i|)^p\right)^{\frac{1}{q}} \left[\left(\sum_{i=1}^{k}|x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{k}|y_i|^p\right)^{\frac{1}{p}}\right],$$

therefore

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

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

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

□

References

[1] S. Lang: Linear Algebra, UTM, Springer, (2nd edition) 1986.

[2] S. Lang Calculus of Several Variables, UTM, Springer, (3d edition) 1987.

[3] M. O'Searcoid: Metric Spaces, Springer, 2007.