# CS724: Topics in Algorithms
# Dissimilarities, Metrics, and Ultrametrics

Prof. Dan A. Simovici

## Definition

A *dissimilarity* on a set $S$ is a function $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ such that $d(x,x) = 0$ and $d(x,y) = d(y,x)$ for every $x, y \in S$. If $d(x,y) = 0$ implies $x = y$, then $d$ is a *definite dissimilarity*.
The set of dissimilarities is denoted by $\mathcal{D}_S$; the set of definite dissimilarities is denoted by $\mathcal{DD}_S$.
The pair $(S, d)$ is a *dissimilarity space*.

If $d$ and $d'$ are dissimilarities on $S$ we write $d \leqslant d'$ if $d(x, y) \leqslant d'(x, y)$ for every $x, y \in S$.
If $d$ is a definite dissimilarity and $d \leqslant d'$, then $d'$ is a definite dissimilarity. Indeed, suppose that $d'(x, y) = 0$. This implies $d(x, y) = 0$. Since $d$ is definite, we have $x = y$, hence $d'$ is definite.
The definition of an *extended dissimilarity* is exactly the same as the definition of a dissimilarity except that the set of values is $\mathbb{R}_{\geqslant 0} \cup \{\infty\}$.
The attribute "extended" will be applied to all types of dissimilarities defined in the sequel if they range over $\mathbb{R}_{\geqslant 0} \cup \{\infty\}$.

A *triangle* in a dissimilarity space $(S, d)$ is a three-element subset $\{x_1, x_2, x_3\}$ of $S$. The triangle $\{x_1, x_2, x_3\}$ is denoted by $x_1 x_2 x_3$.

## Definition

A *quasi-metric* on a set $S$ is a dissimilarity $d \in \mathcal{D}_S$ that satisfies the *triangular condition*

$$d(x, y) \leqslant d(x, z) + d(z, y)$$

for every $x, y, z \in S$.
The set of quasi-metrics on a set is denoted by $\mathcal{S}(S)$.
If $d$ is a definite quasi-metric, we say that then $d$ is a *metric*. The set of metrics on $S$ is denoted by $\mathcal{M}(S)$.

## Definition

A *quasi-ultrametric* is a dissimilarity $d$ on a set $S$ that satisfies the *ultrametric condition*:

$$d(x, y) \leqslant \max\{d(x, z), d(z, y)\}$$

for every $x, y, z \in S$. An *ultrametric* is a definite dissimilarity that satisfies the ultrametric condition.

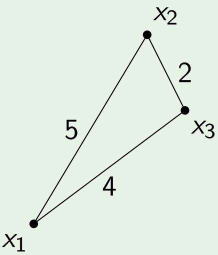
A finite dissimilarity space $(S, d)$, where $S = \{x_1, \ldots, x_n\}$ can be specified by a matrix $D \in (\mathbb{R}_{\geqslant 0})^{n \times n}$, where $D_{ij} = d(x_i, x_j)$ for $1 \leqslant i, j \leqslant n$. Note that $D$ is a symmetrical matrix and the diagonal elements equal 0.

## Example

The matrix $D$

$$D = \begin{pmatrix} 0 & 5 & 4 \\ 5 & 0 & 2 \\ 4 & 2 & 0 \end{pmatrix}$$

specifies the dissimilarity space $(\{x_1, x_2, x_3\}, d)$ shown below.



## Example

Consider the mapping $d : (\mathbf{Seq}_n(S))^2 \longrightarrow \mathbb{R}_{\geqslant 0}$ defined by

$$d(\mathbf{p}, \mathbf{q}) = |\{i \mid 0 \leqslant i \leqslant n-1 \text{ and } \mathbf{p}(i) \neq \mathbf{q}(i)\}|$$

for all sequences $\mathbf{p}, \mathbf{q}$ of length $n$ on the set $S$.
It is easy to see that $d$ is a metric. We justify here only the triangular inequality. Let $\mathbf{p}, \mathbf{q}, \mathbf{r}$ be three sequences of length $n$ on the set $S$. If $\mathbf{p}(i) \neq \mathbf{q}(i)$, then $\mathbf{r}(i)$ must be distinct from at least one of $\mathbf{p}(i)$ and $\mathbf{q}(i)$. Therefore,

$$\begin{aligned} &\{i \mid 0 \leqslant i \leqslant n-1 \text{ and } \mathbf{p}(i) \neq \mathbf{q}(i)\} \\ &\quad \subseteq \quad \{i \mid 0 \leqslant i \leqslant n-1 \text{ and } \mathbf{p}(i) \neq \mathbf{r}(i)\} \\ &\qquad\qquad \cup \{i \mid 0 \leqslant i \leqslant n-1 \text{ and } \mathbf{r}(i) \neq \mathbf{q}(i)\}, \end{aligned}$$

which implies the triangular inequality.

The ultrametric inequality is stronger than the triangular inequality for, if $d(x, y) \leqslant \max\{d(x, z), d(z, y)\}$, then $d(x, y) \leqslant d(x, z) + d(z, y)$ for every $x, y, z \in S$. Thus, every quasi-ultrametric is also a quasi-metric and every ultrametric is also a metric.

We will denote by $\mathcal{U}(S)$ the set of quasi-ultrametrics defined on the set $S$.

If $d$ is a dissimilarity on the set $S$ we refer to the pair $(S, d)$ as a *dissimilarity space*.

If $d$ has other properties (e.g. is a metric) we refer to the pair $(S, d)$ as a space with the corresponding property (a metric space).

## Definition

Let $(S, d)$ and $(S', d')$ be two dissimilarity space. An *isometry* or a *morphism of dissimilarity space* is a function $h : S \longrightarrow S'$ such that $d'(h(x), h(y)) = d(x, y)$ for every $x, y \in S$.
If $h$ is a bijection, we say that $h$ is an *isomorphism* of dissimilarity spaces. Note that in this case $d'(y_1, y_2) = d(h^{-1}(y_1), h^{-1}(y_2)$ for every $y_1, y_2 \in S'$.

## Definition

Let $(S, d)$ and $(S', d')$ be two dissimilarity space. The partitions $\pi \in PART(S)$ and $\sigma \in PART(S')$ are *isomorphic* if there exists a isomorphism $\phi : S \longrightarrow S'$ of the dissimilarity spaces $(S, d)$ and $(S', d')$ such that $x \equiv y(\pi)$ if and only if $\phi(x) \equiv \phi(y)(\sigma)$. This is denoted by $(\pi, d) \sim_C (\pi', d')$.

## Theorem

*If $h : S \longrightarrow S'$ is an isometry between the definite dissimilarity space $(S, d)$ and the dissimilarity space $(S', d')$, then $h$ is an injective mapping.*

**Proof:** Suppose that $h(x) = h(y)$, which implies $d'(h(x), h(y)) = 0$. Thus, $d(x, y) = 0$, hence $x = y$ because $d$ is a definite dissimilarity. We conclude that $h$ is injective.

## Definition

Let $(S, d)$ be a dissimilarity space. A partition $\sigma \in PART(S)$ is *nice* or *locally well-separated* if for every $x, u, v \in S$, $x \equiv u(\sigma)$ and $x \not\equiv v(\sigma)$ implies $d(x, u) \leqslant d(x, v)$.

In other words, for a nice partition of a dissimilarity space no element is closer to an element from a different block than it is to any point from its own block.

## Definition

A partition $\sigma$ of a dissimilarity space $(S, d)$ is *perfect* or *globally well-separated* if of all its in-block dissimilarities are smaller than all of its between-block dissimilarities. In other words, for every $u, v, x, y \in S$, if $u \equiv v(\sigma)$ and $x \not\equiv y(\sigma)$, then $d(u, v) < d(x, y)$.

Clearly, every perfect partition is nice; the converse is not true.

## Example

A bipartition $\pi = \{X, Y\}$ of a set $S$ defines a quasi-ultrametric $\delta_\pi$ on $S$ as

$$\delta_\pi(x, y) = \begin{cases} 1 & \text{if } x \not\equiv y(\pi), \\ 0 & \text{otherwise} \end{cases}$$

for $x, y \in S$. If $d_\pi(x, y) = 0$, then $d_\pi(x, y) \leqslant d_\pi(x, z) + d_\pi(z, y)$ for every $z \in S$. If $d_\pi(x, y) = 1$, it means that $x$ and $y$ belong to distinct blocks of $\pi$. Thus, for any $z$ we have either $x \equiv z(\pi)$ and therefore $z \not\equiv y(\pi)$, or $x \equiv z(\pi)$ and $z \equiv y(\pi)$. In the first case we have $\delta_\pi(x, z) = 0$ and $\delta_\pi(z, y) = 1$; in the second case, $\delta_\pi(x, z) = 1$ and $\delta_\pi(z, y) = 0$. In either case, the ultrametric inequality is satisfied.

## Definition

Let $(S, d)$ be a dissimilarity space. The *open sphere* centered in $x_0$ of radius $r$ is the set

$$B(x_0, r) = \{x \in S \mid d(x_0, x) < r\}.$$

The *closed sphere* centered in $x_0$ of radius $r$ is the set

$$B[x_0, r] = \{x \in S \mid d(x_0, x) \leqslant r\}.$$

## Definition

Let $(S, d)$ be a dissimilarity space. The *diameter* of a subset $U$ of $S$ is

$$diam_d(U) = \sup\{d(x, y) \mid x, y \in U\}.$$

A pair of points $x, y \in S$ is *diametrical* if $d(x, y) = diam_d(S)$.
If the dissimilarity $d$ is clear from context, then the subscript $d$ may be omitted.

If $(S, d)$ is a finite dissimilarity space, the supremum in the definition of the diameter of a set can be replaced by the maximum, that is,

$$diam_d(U) = \max\{d(x, y) \mid x, y \in U\}.$$

A mapping $d : S \times S \longrightarrow \hat{\mathbb{R}}_{\geqslant 0}$ can be extended to the set of subsets of $S$ by defining $d(U, V)$ as

$$d(U, V) = \inf\{d(u, v) \mid u \in U \text{ and } v \in V\} \tag{1}$$

for $U, V \in \mathcal{P}(S)$.
Observe that, even if $d$ is a metric, then its extension is not, in general, a metric on $\mathcal{P}(S)$ because it does not satisfy the triangular inequality.
Instead, we can show that for every $U, V, W$ we have

$$d(U, W) \leqslant d(U, V) + diam(V) + d(V, W).$$

Indeed, by the definition of $d(U, V)$ and $d(V, W)$, for every $\epsilon > 0$, there exist $u \in U$, $v, v' \in V$, and $w \in W$ such that

$$d(U, V) \leqslant d(u, v) \leqslant d(U, V) + \tfrac{\epsilon}{2},$$
$$d(V, W) \leqslant d(v', w) \leqslant d(V, W) + \tfrac{\epsilon}{2}.$$

By the triangular axiom, we have

$$d(u, w) \leqslant d(u, v) + d(v, v') + d(v', w).$$

Hence,

$$d(u, w) \leqslant d(U, V) + diam(V) + d(V, W) + \epsilon,$$

which implies

$$d(U, W) \leqslant d(U, V) + diam(V) + d(V, W) + \epsilon$$

for every $\epsilon > 0$. This yields the needed inequality.

## Definition

Let $(S, d)$ be a metric space. The sets $U, V \in \mathcal{P}(S)$ are *separate* if $d(U, V) > 0$.

We denote the number $d(\{u\}, V) = \inf\{d(u, v) \mid v \in V\}$ by $d(u, V)$. It is clear that $u \in V$ implies $d(u, V) = 0$.

A norm is a real-valued function defined on a linear space that satisfies certain conditions and assigns a positive number to any non-zero vector. Norms are intended to model the "length" of vectors. As we shall see, norms generate metrics on linear spaces.

### Definition

Let $V$ be a real linear space. A *seminorm* on $V$ is a mapping $\nu : V \longrightarrow \mathbb{R}$ that satisfies the following conditions:

- $\nu(x + y) \leqslant \nu(x) + \nu(y)$ (subadditivity), and
- $\nu(ax) = |a|\nu(x)$ (positive homogeneity),

for $x, y \in V$ and $a \in \mathbb{R}$.
If $\nu(x) = 0$ implies $x = 0$, then we say that $\nu$ is a *norm*.

By taking $a = 0$ in the second condition of the definition we have $\nu(0) = 0$ for every seminorm.

## Example

The set of real-valued continuous functions defined on the interval $[-1,1]$ is a real linear space. The addition of two such functions $f, g$, is defined by $(f+g)(x) = f(x) + g(x)$ for $x \in [-1,1]$; the multiplication of $f$ by a scalar $a \in \mathbb{R}$ is $(af)(x) = af(x)$ for $x \in [-1,1]$.
Define $\nu(f) = \sup\{|f(x)| \mid x \in [-1,1]\}$. Since $|f(x)| \leqslant \nu(f)$ and $|g(x)| \leqslant \nu(g)$ for $x \in [-1,1]\}$, it follows that $|(f+g)(x)| \leqslant |f(x)| + |g(x)| \leqslant \nu(f) + \nu(g)$. Thus, $\nu(f+g) \leqslant \nu(f) + \nu(g)$.
We denote $\nu(f)$ by $\| f \|$.

## Theorem

*Let $\nu : V \longrightarrow \mathbb{R}$ be a seminorm on a linear space $V$. We have*

$$\nu(x - y) \geqslant |\nu(x) - \nu(y)|,$$

*for $x, y \in V$.*

**Proof:** We have $\nu(x) \leqslant \nu(x - y) + \nu(y)$, so

$$\nu(x) - \nu(y) \leqslant \nu(x - y).$$

Since $\nu(x - y) = |-1|\nu(y - x) \geqslant \nu(y) - \nu(x)$ we have

$$-(\nu(x) - \nu(y)) \leqslant \nu(x) - \nu(y).$$

## Corollary

*If $\nu : V \longrightarrow \mathbb{R}$ is a seminorm on the linear space $V$, then $\nu(x) \geqslant 0$ for $x \in V$.*

**Proof:** Choose $y = 0$ in the previous inequality; we have $\nu(x) \geqslant |\nu(x)| \geqslant 0$.

Every norm defined on a linear space $V$ generates a metric on the space.

**Theorem**

*Each norm $\nu : V \longrightarrow \mathbb{R}_{\geqslant 0}$ on a real linear space $V$ generates a metric on the set $V$ defined by $d_\nu(x, y) = \nu(x - y)$ for $x, y \in V$.*

# Proof

Note that if $d_\nu(x, y) = \nu(x - y) = 0$, it follows that $x - y = 0$; that is, $x = y$.
The symmetry of $d_\nu$ is obvious and so we need to verify only the triangular axiom. Let $x, y, z \in L$. Applying the subaditivity of norms we have we have

$$\nu(x - z) = \nu(x - y + y - z) \leqslant \nu(x - y) + \nu(y - z)$$

or, equivalently, $d_\nu(x, z) \leqslant d_\nu(x, y) + d_\nu(y, z)$, for every $x, y, z \in L$, which concludes the argument.

We refer to $d_\nu$ as the *metric induced by the norm $\nu$ on the linear space* $V$. Observe that the norm $\nu$ can be expressed using $d_\nu$ as $\nu(x) = d_\nu(x, 0)$ for $x \in V$.

A simple and interesting property of triangles in ultrametric spaces is given next.

**Theorem**

*Let $(S, d)$ be an ultrametric space and let $t = xyz$ be a triangle in $(S, d)$. The two sides of $t$ which are not the smallest have equal length.*

**Proof:** If $d(x, y)$ is the smallest of the numbers $d(x, y), d(y, z), d(x, z)$ we have

$$\begin{aligned} d(y, z) &\leqslant \max\{d(x, y), d(x, z)\} = d(x, z), \\ d(x, z) &\leqslant \max\{d(x, y), d(y, z)\} = d(y, z), \end{aligned}$$

so $d(x, z) = d(y, z)$.

In an ultrametric space any triangle is isosceles and the side that is not equal to the two others cannot be longer than these.

**Theorem**

*Let $S = \{x_1, \ldots, x_m\}$ be a finite set with $m \geqslant 2$ such that $(S, d)$ is a quasi-ultrametric space. The range of $d$, $Ran(d)$ contains at most $m$ distinct values.*

# Proof

The proof is by induction on $m$. The basis step, $m = 1$ is immediate because $\text{Ran}(d) = 0$. Suppose that the statement holds for sets of size less than $m$, let $S = \{x_1, \ldots, x_m\}$ be a set of size $m$ and let $T = \{x_1, \ldots, x_{m-1}\}$. The restriction of $d$ to $T \times T$ contains at most $m - 1$ values. Let $x_k$ be one of the closest element of $T$ to $x_m$. Then, for every element $x_j$ of $T$ we have $d(x_j, x_m) \geqslant d(x_k, x_m)$. Since the triangle $x_j x_k x_m$ is isosceles we may have either

$$d(x_j, x_m) = d(x_k, x_m) \geqslant d(x_j, x_k),$$

or

$$d(x_j, x_k) = d(x_j, x_m) > d(x_k, x_m).$$

In the first case, only one element, $d(x_k, x_m)$ is added to the range of $d$; in the second case, $d(x_j, x_m)$ equals one of the previous values and the size of the range of $d$ remains the same. In either case, the size of the range of $d$ is no larger than $m$.

## Corollary

*For an ultrametric space $(S, d)$ with $|S| = m$, there are at most $m - 1$ non-zero values of $d$.*

## Theorem

**(Egocentricity of Closed Spheres)** *Let $B[x, r]$ be a closed sphere in a finite ultrametric space $(S, d)$. If $y \in B[x, r]$, then $B[x, r] = B[y, r]$. In other words, in a finite ultrametric space, a closed sphere has all its points as centers.*

# Proof

Let $z \in B[x,r]$. Since $d(x,z) \leqslant r$, it follows that

$$d(y,z) \leqslant \max\{d(y,x), d(x,z)\} \leqslant r,$$

so $z \in B[y,r]$, which implies $B[x,r] \subseteq B[y,r]$.
If $u \in B[y,r]$ we have $d(y,u) \leqslant r$. This implies

$$d(x,u) \leqslant \max\{d(x,y), d(y,u)\} \leqslant r,$$

hence $u \in B[x,r]$. Thus, $B[y,r] \subseteq B[x,r]$. We conclude that $B[y,r] = B[x,r]$.

## Corollary

*If two closed spheres* $B[x, r]$ *and* $B[y, r']$ *in a finite ultrametric space have a point in common, then one of the closed spheres is included in the other.*

## Theorem

**(Egocentricity of Open Spheres)** *Let* $B(x, r)$ *be an open sphere in a finite ultrametric space* $(S, d)$. *If* $y \in B(x, r)$, *then* $B(x, r) = B(y, r)$. *In other words, in a finite ultrametric space, an open sphere has all its points as centers.*

# Proof

Since $y \in B(x,r)$, we have $d(x,y) < r$. For $z \in B(x,r)$ we have $d(x,z) < r$.
Since the triangle $xyz$ is isosceles the following three cases may occur:

- $d(x,y) \leqslant d(x,z) = d(y,z) < r$ (because $d(x,z) < r$), or
- $d(x,z) \leqslant d(x,y) = d(y,z) < r$ (because $d(x,y) < r$), or
- $d(y,z) \leqslant d(x,y) = d(x,z) < r$ (because $d(x,y) < r$).

Thus, in every case $d(y,z) < r$, so $z \in B(y,r)$. Thus, $B(x,r) \subseteq B(y,r)$.
Conversely, suppose that $z \in B(y,r)$, that is, $d(y,z) < r$. Since $d(y,x) < r$ the following cases may occur

- $d(x,y) \leqslant d(x,z) = d(y,z) < r$ (because $d(y,z) < r$), or
- $d(x,z) \leqslant d(x,y) = d(y,z) < r$ (because $d(y,z) < r$), or
- $d(y,z) \leqslant d(x,y) = d(x,z) < r$ (because $d(y,x) < r$).

Therefore, in every case $d(x,z) < r$, so $B(y,r) \leqslant B(x,r)$. We conclude that $B(x,r) = B(y,z)$.

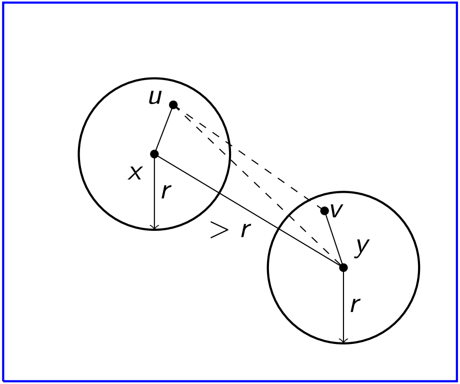
**Theorem**

*Let $B[x, r]$ and $B[y, r]$ be two disjoint closed spheres in the finite ultrametric space $(S, d)$ having the same radius $r$. Then, for every $u \in B[x, r]$ and $v \in B[y, r]$ we have $d(u, v) = d(x, y)$.*

UMASS BOSTON

# Proof

Let $k = d(u, v)$. Since $B[x, r]$ and $B[y, r]$ are disjoint closed spheres, it follows that $k > r$. Since $xyu$ is an isosceles triangle, it follows that $d(u, y) = k$. Since $d(v, y) \leqslant r$, we have $d(v, u) = d(u, y) = k$.



## Theorem

*Let $(S, d)$ be a finite ultrametric space. If $B[x_0, r]$ is a closed sphere in $(S, d)$ such that $B[x_0, r] \neq \{x_0\}$, then $diam_d(B[x_0, r]) = r$.*

**Proof:** Since $B[x_0, r]$ contains a point $x$ distinct from $x_0$ and $d(x_0, x) \leqslant r$, it is clear that the diameter of $B[x_0, r]$ is at least equal to $r$.
Let $u, v$ be elements of $B[x_0, r]$ such that $d(u, v) = diam(B[x_0, r])$. Since $d(x_0, u) \leqslant r$ and $d(x_0, v) \leqslant r$, and $d(u, v)$ is one of the largest sides of the triangle $x_0 uv$, it follows that we have either $d(u, v) = d(x_0, u)$ or $d(u, v) = d(x_0, v)$. In either case, $d(u, v) \leqslant r$, hence $diam_d(B[x_0, r]) \leqslant r$.
Thus, $diam_d(B[x_0, r]) = r$.

The next statement provides a characterization of set diameters in ultrametric spaces.

### Theorem

*Let $S$ be a finite non-empty set. A function $\tau : \mathcal{P}(X) \longrightarrow \mathbb{R}_{\geqslant 0}$ satisfies the conditions*

- $\tau(A) = 0$ *if and only if $A$ is a singleton, and*
- $\tau(A \cup B) \leqslant \max\{\tau(A \cup C), \tau(C \cup B)\}$

*for all $A, B, C \in \mathcal{P}(S)$ if and only if there exists an ultrametric $d$ on $S$ such that $\tau(A) = diam_d(A)$ for every $A \in \mathcal{P}(S)$.*

# Proof

Suppose that $\tau$ satisfies the conditions of the theorem. Choosing $B = A$ in the second condition implies $\tau(A) \leqslant \tau(A \cup C)$. Thus, if $A \subseteq C$, we have $\tau(A) \leqslant \tau(C)$, so $\tau$ is isotonic.
Let $c \in C$. Choosing $A = \{c\}$, it follows that $0 = \tau(c) \leqslant \tau(C)$, hence $\tau(C) \geqslant 0$, so $\tau$ is non-negative.
Define the function $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ as $d(x, y) = \tau(\{x, y\})$. Property (i) implies that $d(x, y) = 0$ if and only if $x = y$; moreover, $d$ is non-negative.

Choosing $A = \{a\}$, $B = \{b\}$, and $C = \{c\}$ in the second condition implies $d(a, b) \leqslant \max\{d(a, c), d(c, b)\}$, so $d$ is an ultrametric.
It remains to show that $\tau(X) = diam_d(X)$ for $|X| \geqslant 3$. Let $X = \{x_1, \ldots, x_n\}$. For $1 \leqslant j \leqslant n$ prove that $\tau(X) \leqslant \max_{i,j} d(x_i, x_j)$ for $1 \leqslant i, j \leqslant n$. The argument is by induction on $n$ and we can write:

$$\begin{aligned}
\tau(X) &= \tau(\{x_1, \ldots, x_{n-1}\} \cup \{x_n\}) \\
&\quad \text{(setting } A = \{x_1, \ldots, x_{n-1}\} \text{ and } B = \{x_n\}) \\
&\leqslant \max\{\tau(\{x_1, \ldots, x_{n-1}\} \cup \{x_j\}), \tau(\{x_n, x_j\})\} \\
&\quad \text{(setting } C = \{x_j\}) \\
&= \max\{\tau(\{x_1, \ldots, x_{n-1}\}), d(\{x_n, x_j\})\} \\
&\leqslant \max_{i,j} d(x_i, x_j).
\end{aligned}$$

# Proof (cont'd)

By repeating this process we obtain

$$\tau(X) \leqslant \max\{d(x_i, x_j) \mid 1 \leqslant i, j \leqslant n\} = diam_d(X).$$

Since $\tau$ is an isotonic mapping, if $\{x_i, x_j\}$ is a diametrical pair of points in $X$, then

$$\tau(X) \geqslant \tau(\{x_i, x_j\}) = d(x_i, x_j) = diam_d(X),$$

hence $\tau(X) = diam_d(X)$.

Conversely, suppose that $d$ is an ultrametric on $X$ such that $\tau(A) = diam_d(A)$ for $A \in \mathcal{F}(X)$. This implies that if $\tau(A) = 0$, then $A$ is a singleton.
Let $a, b$ be two diametrical points in $A \cup B$. There are three cases to consider in order to prove (ii):
If both $a$ and $b$ belong to $A$, then

$$\max\{d(x,y) \mid x,y \in A \cup B\} = \max\{d(x,y) \mid x,y \in A\}.$$

Therefore,

$$\begin{aligned} diam_d(A \cup B) &= \max\{d(x,y) \mid x,y \in A\} \\ &\leqslant \max\{d(x,z) \mid x,z \in A \cup C\} \\ &\leqslant \max\{\max\{d(x,z) \mid x,z \in A \cup C\}, \\ &\qquad \max\{d(y,z) \mid y,z \in B \cup C\}\} \\ &= \max\{diam_d(A \cup C), diam_d(C \cup B)\}, \end{aligned}$$

hence $\tau(A \cup B) \leqslant \max\{\tau(A \cup C), \tau(C \cup B)\}$.

If both $a$ and $b$ belong to $B$, then we obtain

$$\max\{d(x,y) \mid x,y \in A \cup B\} = \max\{d(x,y) \mid x,y \in B\},$$

as above, which implies $\tau(A \cup B) \leqslant \max\{\tau(A \cup C), \tau(C \cup B)\}$.

If $a \in A$ and $b \in B$, then

$$\begin{aligned} \tau(A \cup B) &= \max\{d(x,y) \mid x, y \in A \cup B\} \\ &= \max\{d(x,y) \mid x \in A, y \in B\} = d(a,b), \end{aligned}$$

where $a \in A$ and $b \in B$. Then, for each $c \in C$, we have

$$\begin{aligned} diam_d(A \cup B) &= \max\{d(x,y) \mid x \in A, y \in B\} \\ &= d(a,b) \leqslant \max\{d(a,c), d(b,c)\} \\ &\leqslant \max\{\max\{d(x,z) \mid x \in A, z \in C\}, \\ &\quad \max\{d(y,z) \mid y \in B, z \in C\}\} \\ &\leqslant \max\{diam_d(A \cup C), diam_d(C \cup B)\}. \end{aligned}$$

## Theorem

*Let $(S, d)$ be a finite ultrametric space. For every $r \geqslant 0$ the family of spheres $\sigma_r = \{B[x, r] \mid x \in S\}$ is a partition of $S$.*

**Proof:** Since $x \in B[x, r]$ for every $x \in S$, we have $\bigcup_{x \in S} B[x, r] = S$. Suppose that $B[x, r] \cap B[y, r] \neq \emptyset$. One of these spheres is included in the other, which implies that they are equal. Thus, two distinct spheres of the same radius are disjoint, which implies that they constitute a partition of $S$.

Let $\mathrm{CLS}(S, d)$ be the collection of subsets that consists of $\emptyset$ and all closed spheres of the ultrametric space $(S, d)$.

**Theorem**

*If $(S, d)$ is a finite ultrametrc space with $|S| = n$, then $|CLS(S, d)| \leqslant 2\,|S|$.*

# Proof

The proof is by strong induction on $n = |S|$. If $n = 1$, $S = \{x\}$ is a singleton and the unique closed sphere is $\{a\}$. Thus, $|\text{CLS}(S, d)| = 2$ and the inequality is satisfied.

Suppose that the inequality holds for any dissimilarity space $(S, d)$ with $|S| \leqslant n$.

Let now $(S, d)$ be an ultrametric space with $|S| = n + 1$, $S = \{x_0, \ldots, x_{n-1}, x_n\}$. Suppose that

$$\begin{aligned} & \min\{d(x_i, x_j) \mid x_i \neq x_j\} \\ = \ & d(x_0, x_1) \leqslant \cdots \leqslant d(x_0, x_{m-1}) < d(x_0, x_m) \\ = \ & \cdots = d(x_0, x_{n-1}) = d(x_0, x_n) = \max\{d(x_i, x_j) \mid x_i \neq x_j\}, \end{aligned}$$

and let

$$\begin{aligned} S_{m-1} &= \{x_0, x_1, \ldots, x_{m-1}\}, \\ S_m &= \{x_m, x_{m+1}, \ldots, x_n\}. \end{aligned}$$

# Proof (cont'd)

By the inductive assumption, $|\text{CLS}(S_{m-1}, d)| \leqslant 2|S_{m-1}| = 2m$ and $|\text{CLS}(S_m, d)| \leqslant 2|S_m| = 2(n - m + 1)$.
All points $x_k$ for $k \geqslant m$ are at the same distance $d(x_0, x_n)$ from all points $x_j$ with $j < m$. Thus, $\text{CLS}(S, d)) \subseteq \text{CLS}(S_{m-1}, d) \cup \text{CLS}(S_m, d) \cup \{S\}$.
Since the empty set belongs to both $\text{CLS}(S_{m-1}, d)$ and $\text{CLS}(S_m, d)$, $\text{CLS}(S, d) \leqslant 2m + 2(n - m + 1) + 1 - 1 = 2(n + 1) = 2|S|$.

For an ultrametric space $(S, d)$ the set $\mathrm{CLS}(S, d)$ is a lattice.
In other words, for two closed spheres $B, B'$, their meet is just their intersection and is just the smaller of the two closed spheres. The join of $B, B'$ is either the larger of the closed spheres (if one is included in the other), or the ball of radius $r = d(x, y)$, where $x \in B$ and $y \in B'$, having the center at any point of $B \cup B'$.

### Theorem

*Let $d$ be a dissimilarity on a set $S$ and let $U_d$ be the set of quasi-ultrametrics $U_d = \{e \in \mathcal{U}_S \mid e \leqslant d\}$. The set $U_d$ has a largest element in the poset $(\mathcal{D}\mathcal{D}_S, \leqslant)$.*

The set $U_d$ is nonempty because the zero dissimilarity $d_0$ given by $d_0(x, y) = 0$ for every $x, y \in S$ is a quasi-ultrametric and $d_0 \leqslant d$. Since the set $\{e(x, y) \mid e \in U_d\}$ has $d(x, y)$ as an upper bound, it is possible to define the mapping $e_1 : S^2 \longrightarrow \mathbb{R}_{\geq 0}$ as $e_1(x, y) = \sup\{e(x, y) \mid e \in U_d\}$ for $x, y \in S$. It is clear that $e \leqslant e_1$ for every quasi-ultrametric $e$. We claim that $e_1$ is an ultrametric on $S$.

# Proof (cont'd)

We prove only that $e_1$ satisfies the ultrametric inequality. Suppose that there exist $x, y, z \in S$ such that $e_1$ violates the ultrametric inequality; that is,

$$\max\{e_1(x,z), e_1(z,y)\} < e_1(x,y).$$

This is equivalent to

$$\begin{aligned} & \sup\{e(x,y) \mid e \in U_d\} \\ & \quad > \max\{\sup\{e(x,z) \mid e \in U_d\}, \sup\{e(z,y) \mid e \in U_d\}\}. \end{aligned}$$

Thus, there exists $\hat{e} \in U_d$ such that

$$\hat{e}(x,y) > \sup\{e(x,z) \mid e \in U_d\}, \text{ and } \hat{e}(x,y) > \sup\{e(z,y) \mid e \in U_d\}.$$

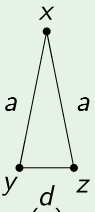
In particular, $\hat{e}(x,y) > \hat{e}(x,z)$ and $\hat{e}(x,y) > \hat{e}(z,y)$, which contradicts the fact that $\hat{e}$ is an quasi-ultrametric.

The quasi-ultrametric defined above is known as the *subdominant quasi-ultrametric for the dissimilarity $d$*.

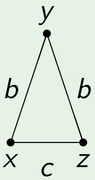
In general, the infimum of a set of quasi-ultrametrics is not necessarily an ultrametric.

## Example

Consider a three-element set $S = \{x, y, z\}$, four distinct nonnegative numbers $a, b, c, d$ such that $a > b > c > d$ and the quasi-ultrametrics $d$ and $d'$ defined by the triangles shown in Figures (a) and (b), respectively.




(a)




(b)

The dissimilarity $d_0$ defined by $d_0(u, v) = \min\{d(u, v), d'(u, v)\}$ for $u, v \in S$ is given by

$$d_0(x, y) = b, d_0(y, z) = d, \text{ and } d_0(x, z) = c,$$

and $d_0$ is clearly not a quasi-ultrametric because the triangle $xyz$ is not isosceles.

The inequalities developed in this section are essential for the study of norms and metrics in $\mathbb{R}^n$.

**Lemma**

*Let $p, q \in \mathbb{R} - \{0, 1\}$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then we have $p > 1$ if and only if $q > 1$. Furthermore, one of the numbers $p, q$ belongs to the interval $(0, 1)$ if and only if the other number is negative.*

## Lemma

*Let* $p, q \in \mathbb{R} - \{0, 1\}$ *be two numbers such that* $\frac{1}{p} + \frac{1}{q} = 1$ *and* $p > 1$*. Then, for every* $a, b \in \mathbb{R}_{\geqslant 0}$*, we have*

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q},$$

*where the equality holds if and only if* $a = b^{-\frac{1}{1-p}}$*.*

# Proof

We have $q > 1$. Consider the function $f(x) = \frac{x^p}{p} + \frac{1}{q} - x$ for $x \geqslant 0$. We have $f'(x) = x^{p-1} - 1$, so the minimum is achieved when $x = 1$ and $f(1) = 0$. Thus,

$$f\left(ab^{-\frac{1}{p-1}}\right) \geqslant f(1) = 0,$$

which amounts to

$$\frac{a^p b^{-\frac{p}{p-1}}}{p} + \frac{1}{q} - ab^{-\frac{1}{p-1}} \geqslant 0.$$

By multiplying both sides of this inequality by $b^{\frac{p}{p-1}}$, we obtain the desired inequality.

Observe that if $\frac{1}{p} + \frac{1}{q} = 1$ and $p < 1$, then $q < 0$. In this case, we have the reverse inequality

$$ab \geqslant \frac{a^p}{p} + \frac{b^q}{q}. \tag{2}$$

which can be shown by observing that the function $f$ has a maximum in $x = 1$. The same inequality holds when $q < 1$ and therefore $p < 0$.

## Theorem

**The Hölder Inequality** *Let* $a_1, \ldots, a_n$ *and* $b_1, \ldots, b_n$ *be* $2n$ *nonnegative numbers, and let* $p$ *and* $q$ *be two numbers such that* $\frac{1}{p} + \frac{1}{q} = 1$ *and* $p > 1$. *We have:*

$$\sum_{i=1}^{n} a_i b_i \leq \left( \sum_{i=1}^{n} a_i^p \right)^{\frac{1}{p}} \cdot \left( \sum_{i=1}^{n} b_i^q \right)^{\frac{1}{q}}.$$

# Proof

If $a_1 = \cdots = a_n = 0$ or if $b_1 = \cdots = b_n = 0$, then the inequality is clearly satisfied. Therefore, we may assume that at least one of $a_1, \ldots, a_n$ and at least one of $b_1, \ldots, b_n$ is non-zero. Define the numbers

$$x_i = \frac{a_i}{\left(\sum_{i=1}^n a_i^p\right)^{\frac{1}{p}}} \text{ and } y_i = \frac{b_i}{\left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}}$$

for $1 \leqslant i \leqslant n$. A previous Lemma applied to $x_i, y_i$ yields

$$\frac{a_i b_i}{\left(\sum_{i=1}^n a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}} \leqslant \frac{1}{p} \frac{a_i^p}{\sum_{i=1}^n a_i^p} + \frac{1}{q} \frac{b_i^p}{\sum_{i=1}^n b_i^p}.$$

# Proof (cont'd)

Adding these inequalities, we obtain

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}$$

because $\frac{1}{p} + \frac{1}{q} = 1$.

The nonnegativity of the numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$ can be relaxed by using absolute values.

## Theorem

*Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ numbers and let $p$ and $q$ be two numbers such that $\frac{1}{p} + \frac{1}{q} = 1$ and $p > 1$. We have*

$$\left| \sum_{i=1}^{n} a_i b_i \right| \leq \left( \sum_{i=1}^{n} |a_i|^p \right)^{\frac{1}{p}} \cdot \left( \sum_{i=1}^{n} |b_i|^q \right)^{\frac{1}{q}}.$$

## Corollary

**(The Cauchy-Schwarz Inequality for $\mathbb{R}^n$)** *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ real numbers. We have*

$$\left| \sum_{i=1}^{n} a_i b_i \right| \leq \sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2}.$$

## Theorem

**(Minkowski's Inequality)** *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ nonnegative real numbers. If $p \geqslant 1$, we have*

$$\left(\sum_{i=1}^{n}(a_i + b_i)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}}.$$

*If $p < 1$, the inequality sign is reversed.*

# Proof

For $p = 1$, the inequality is immediate. Therefore, we can assume that $p > 1$. Note that

$$\sum_{i=1}^{n}(a_i + b_i)^p = \sum_{i=1}^{n} a_i(a_i + b_i)^{p-1} + \sum_{i=1}^{n} b_i(a_i + b_i)^{p-1}.$$

By Hölder's inequality for $p, q$ such that $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\begin{aligned}\sum_{i=1}^{n} a_i(a_i + b_i)^{p-1} &\leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n}(a_i + b_i)^{(p-1)q}\right)^{\frac{1}{q}} \\ &= \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n}(a_i + b_i)^{p}\right)^{\frac{1}{q}}.\end{aligned}$$

# Proof (cont'd)

Similarly, we can write

$$\sum_{i=1}^{n} b_i(a_i+b_i)^{p-1} \leq \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{1}{q}}.$$

Adding the last two inequalities yields

$$\sum_{i=1}^{n}(a_i+b_i)^p \leq \left(\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}}\right) \left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{1}{q}},$$

which is equivalent to inequality

$$\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}}.$$

## Example

For $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ the *Euclidean metric* is the mapping

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

To prove the triangular inequality, let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. Choosing $a_i = x_i - y_i$ and $b_i = y_i - z_i$ for $1 \leqslant i \leqslant n$ in Minkowski's inequality implies

$$\sqrt{\sum_{i=1}^{n} (x_i - z_i)^2} \leq \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} + \sqrt{\sum_{i=1}^{n} (y_i - z_i)^2},$$

which amounts to $d(\mathbf{x}, \mathbf{z}) \leqslant d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$. Thus, we conclude that $d$ is indeed a metric on $\mathbb{R}^n$.

### Theorem

*For* $p \geqslant 1$*, the function* $\nu_p : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geqslant 0}$ *defined by*

$$\nu_p(x_1, \ldots, x_n) = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}},$$

*where* $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$*, is a norm on* $\mathbb{R}^n$*.*

# Proof

Let $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Minkowski's inequality applied to the nonnegative numbers $a_i = |x_i|$ and $b_i = |y_i|$ amounts to

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

Since $|x_i + y_i| \leqslant |x_i| + |y_i|$ for every $i$, we have

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

that is, $\nu_p(\mathbf{x} + \mathbf{y}) \leqslant \nu_p(\mathbf{x}) + \nu_p(\mathbf{y})$.
Thus, $\nu_p$ is a norm on $\mathbb{R}^n$.

## Example

The mapping $\nu_1 : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by

$$\nu_1(\mathbf{x}) = |x_1| + |x_2| + \cdots + |x_n|,$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is a norm on $\mathbb{R}^n$.

## Example

A special norm on $\mathbb{R}^n$ is the function $\nu_\infty : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geqslant 0}$ given by

$$\nu_\infty(\mathbf{x}) = \max\{|x_i| \mid 1 \leqslant i \leqslant n\}$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.
We start from the inequality

$$|x_i + y_i| \leqslant |x_i| + |y_i| \leq \nu_\infty(\mathbf{x}) + \nu_\infty(\mathbf{y})$$

for every $i$, $1 \leqslant i \leqslant n$. This in turn implies

$$\nu_\infty(\mathbf{x} + \mathbf{y}) = \max\{|x_i + y_i| \mid 1 \leqslant i \leqslant n\} \leqslant \nu_\infty(\mathbf{x}) + \nu_\infty(\mathbf{y}),$$

which gives the desired inequality.

$\nu_\infty$ can be regarded as a limit case of the norms $\nu_p$. Indeed, let $\mathbf{x} \in \mathbb{R}^n$ and let $M = \max\{|x_i| \mid 1 \leq i \leqslant n\} = |x_{\ell_1}| = \cdots = |x_{\ell_k}|$ for some $\ell_1, \ldots, \ell_k$, where $1 \leqslant \ell_1, \ldots, \ell_k \leqslant n$. Here $x_{\ell_1}, \ldots, x_{\ell_k}$ are the components of $\mathbf{x}$ that have the maximal absolute value and $k \geqslant 1$. We can write

$$\lim_{p\to\infty} \nu_p(\mathbf{x}) = \lim_{p\to\infty} M \left( \sum_{i=1}^{n} \left( \frac{|x_i|}{M} \right)^p \right)^{\frac{1}{p}} = \lim_{p\to\infty} M(k)^{\frac{1}{p}} = M,$$

which justifies the notation $\nu_\infty$.

We use the alternative notation $\| \mathbf{x} \|_p$ for $\nu_p(\mathbf{x})$. We refer $\| \mathbf{x} \|_2$ as the *Euclidean norm* of $\mathbf{x}$ and we denote this norm simply by $\| \mathbf{x} \|$ when there is no risk of confusion.

## Example

For $p \geqslant 1$, let $\ell_p$ be the set that consists of sequences of real numbers $\mathbf{x} = (x_0, x_1, \ldots)$ such that the series $\sum_{i=0}^{\infty} |x_i|^p$ is convergent. We can show that $\ell_p$ is a linear space.

Let $\mathbf{x}, \mathbf{y} \in \ell_p$ be two sequences in $\ell_p$. Using Minkowski's inequality we have

$$\sum_{i=0}^{n} |x_i + y_i|^p \leqslant \sum_{i=0}^{n} (|x_i| + |y_i|)^p \leq \sum_{i=0}^{n} |x_i|^p + \sum_{i=0}^{n} |y_i|^p,$$

which shows that $\mathbf{x} + \mathbf{y} \in \ell_p$. It is immediate that $\mathbf{x} \in \ell_p$ implies $a\mathbf{x} \in \ell_p$ for every $a \in \mathbb{R}$ and $\mathbf{x} \in \ell_p$.

For $p \geqslant 1$, then $d_p$ denotes the metric $d_{\nu_p}$ induced by the norm $\nu_p$ on the linear space $\mathbb{R}^n$ known as the *Minkowski metric* on $\mathbb{R}^n$.
If $p = 2$, we have the *Euclidean metric* on $\mathbb{R}^n$ given by

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

For $p = 1$, we have

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|.$$

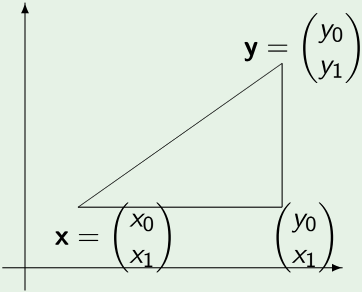
This metric is known also as the *city-block metric*.
The norm $\nu_\infty$ generates the metric $d_\infty$ given by

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max\{|x_i - y_i| \mid 1 \leqslant i \leqslant n\},$$

also known as the *Chebyshev metric*.

## Example

In the special case of $\mathbb{R}^2$ for $\mathbf{x} = (x_0, x_1)$ and $\mathbf{y} = (y_0, y_1)$, then $d_2(\mathbf{x}, \mathbf{y})$ is the length of the hypotenuse of the right triangle and $d_1(\mathbf{x}, \mathbf{y})$ is the sum of the lengths of the two legs of the triangle.



We can compare the norms $\nu_p$ (and the metrics of the form $d_p$) that were introduced on $\mathbb{R}^n$. We begin with a preliminary result.

### Lemma

*Let $a_1, \ldots, a_n$ be $n$ positive numbers. If $p$ and $q$ are two positive numbers such that $p \leqslant q$, then*

$$\left(a_1^p + \cdots + a_n^p\right)^{\frac{1}{p}} \geq \left(a_1^q + \cdots + a_n^q\right)^{\frac{1}{q}}.$$

# Proof

Let $f : \mathbb{R}^{>0} \longrightarrow \mathbb{R}$ be the function defined by

$$f(r) = (a_1^r + \cdots + a_n^r)^{\frac{1}{r}} .$$

Since

$$\ln f(r) = \frac{\ln (a_1^r + \cdots + a_n^r)}{r},$$

it follows that

$$\frac{f'(r)}{f(r)} = -\frac{1}{r^2}(a_1^r + \cdots + a_n^r) + \frac{1}{r} \cdot \frac{a_1^r \ln a_1 + \cdots + a_n^r \ln a_r}{a_1^r + \cdots + a_n^r}.$$

To prove that $f'(r) < 0$, it suffices to show that

$$\frac{a_1^r \ln a_1 + \cdots + a_n^r \ln a_r}{a_1^r + \cdots + a_n^r} \leqslant \frac{\ln (a_1^r + \cdots + a_n^r)}{r}.$$

This last inequality is easily seen to be equivalent to

$$\sum_{i=1}^{n} \frac{a_i^r}{a_1^r + \cdots + a_n^r} \ln \frac{a_i^r}{a_1^r + \cdots + a_n^r} \leqslant 0,$$

which holds because

### Theorem

*Let $p$ and $q$ be two positive numbers such that $p \leqslant q$. For every $\mathbf{u} \in \mathbb{R}^n$, we have $\| \mathbf{u} \|_p \geq \| \mathbf{u} \|_q$.*

**Proof:** This statement follows immediately from previous Lemma.

## Corollary

*Let $p, q$ be two positive numbers such that $p \leqslant q$. For every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have $d_p(\mathbf{x}, \mathbf{y}) \geq d_q(\mathbf{x}, \mathbf{y})$.*

## Example

For $p = 1$ and $q = 2$ the inequality of a previous theorem becomes

$$\sum_{i=1}^{n} |u_i| \leqslant \sqrt{\sum_{i=1}^{n} |u_i|^2},$$

which is equivalent to

$$\frac{\sum_{i=1}^{n} |u_i|}{n} \leqslant \sqrt{\frac{\sum_{i=1}^{n} |u_i|^2}{n}}. \tag{3}$$

## Theorem

*Let $p \geqslant 1$. For every $\mathbf{x} \in \mathbb{R}^n$ we have*

$$\| \mathbf{x} \|_\infty \leqslant \| \mathbf{x} \|_p \leqslant n \| \mathbf{x} \|_\infty .$$

**Proof:** Starting from the definition of $\nu_p$ we have

$$\| \mathbf{x} \|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \leq n^{\frac{1}{p}} \max_{1 \leqslant i \leq n} |x_i| = n^{\frac{1}{p}} \| \mathbf{x} \|_\infty .$$

The first inequality is immediate.

## Corollary

*Let p and q be two numbers such that* $p, q \geqslant 1$. *There exist two constants* $c, d \in \mathbb{R}_{>0}$ *such that*

$$c \parallel \mathbf{x} \parallel_q \leqslant \parallel \mathbf{x} \parallel_p \leqslant d \parallel \mathbf{x} \parallel_q$$

*for* $\mathbf{x} \in \mathbb{R}^n$.

**Proof:** Since $\parallel \mathbf{x} \parallel_\infty \leqslant \parallel \mathbf{x} \parallel_p$ and $\parallel x \parallel_q \leqslant n \parallel \mathbf{x} \parallel_\infty$, it follows that $\parallel x \parallel_q \leqslant n \parallel \mathbf{x} \parallel_p$. Exchanging the roles of $p$ and $q$, we have $\parallel x \parallel_p \leqslant n \parallel \mathbf{x} \parallel_q$, so

$$\frac{1}{n} \parallel \mathbf{x} \parallel_q \leqslant \parallel \mathbf{x} \parallel_p \leqslant n \parallel \mathbf{x} \parallel_q$$

for every $\mathbf{x} \in \mathbb{R}^n$.

## Corollary

*For every* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *and* $p \geqslant 1$, *we have* $d_\infty(\mathbf{x}, \mathbf{y}) \leqslant d_p(\mathbf{x}, \mathbf{y}) \leqslant n d_\infty(\mathbf{x}, \mathbf{y})$. *Further, for* $p, q > 1$, *there exist* $c, d \in \mathbb{R}_{>0}$ *such that*

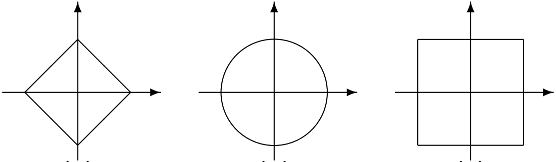
$$c d_q(\mathbf{x}, \mathbf{y}) \leqslant d_p(\mathbf{x}, \mathbf{y}) \leqslant c d_q(\mathbf{x}, \mathbf{y})$$

*for* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

If $p \leqslant q$, then the closed sphere $B_{d_p}[\mathbf{x}, r]$ is included in the closed sphere $B_{d_q}[\mathbf{x}, r]$. For example, we have

$$B_{d_1}[\mathbf{0}, 1] \subseteq B_{d_2}[\mathbf{0}, 1] \subseteq B_{d_\infty}[\mathbf{0}, 1].$$

In Figures (a) - (c) we represent the closed spheres $B_{d_1}[\mathbf{0}, 1]$, $B_{d_2}[\mathbf{0}, 1]$, and $B_{d_\infty}[\mathbf{0}, 1]$.



(a) (b) (c)

## Theorem

*Let* $x_1, \ldots, x_m$ *and* $y_1, \ldots, y_m$ *be* $2m$ *nonnegative numbers such that* $\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i = 1$ *and let* $p$ *and* $q$ *be two positive numbers such that* $\frac{1}{p} + \frac{1}{q} = 1$. *We have*

$$\sum_{j=1}^{m} x_j^{\frac{1}{p}} y_j^{\frac{1}{q}} \leqslant 1.$$

**Proof:** The Hölder inequality applied to $x_1^{\frac{1}{p}}, \ldots, x_m^{\frac{1}{p}}$ and $y_1^{\frac{1}{q}}, \ldots, y_m^{\frac{1}{q}}$ yields the needed inequality

$$\sum_{j=1}^{m} x_j^{\frac{1}{p}} y_j^{\frac{1}{q}} \leq \sum_{j=1}^{m} x_j \sum_{j=1}^{m} y_j = 1$$

UMASS BOSTON

We can formulate now a generalization of the Hölder Inequality.

### Theorem

*Let $A$ be an $n \times m$ matrix, $A = (a_{ij})$, having positive entries such that $\sum_{j=1}^{m} a_{ij} = 1$ for $1 \leqslant i \leq n$. If $\mathbf{p} = (p_1, \ldots, p_n)$ is an $n$-tuple of positive numbers such that $\sum_{i=1}^{n} p_i = 1$, then*

$$\sum_{j=1}^{m} \prod_{i=1}^{n} a_{ij}^{p_i} \leqslant 1.$$

# Proof

The argument is by induction on $n \geqslant 2$. The basis case, $n = 2$ follows immediately by choosing $p = \frac{1}{p_1}$, $q = \frac{1}{p_2}$, $x_j = a_{1j}$, and $y_j = a_{2j}$ for $1 \leqslant j \leqslant m$.

Suppose that the statement holds for $n$, let $A$ be an $(n+1) \times m$-matrix having positive entries such that $\sum_{j=1}^{m} a_{ij} = 1$ for $1 \leqslant i \leqslant n+1$, and let $\mathbf{p} = (p_1, \ldots, p_n, p_{n+1})$ be such that $p_1 + \cdots + p_n + p_{n+1} = 1$.

It is easy to see that

$$\sum_{j=1}^{m} \prod_{i=1}^{n+1} a_{ij}^{p_i} \leq \sum_{j=1}^{m} a_{1j}^{p_1} a_{n-1\,j}^{p_{n-1}} (a_{nj} + a_{n+1\,j})^{p_n + p_{n+1}}.$$

By applying the inductive hypothesis, we have $\sum_{j=1}^{m} \prod_{i=1}^{n+1} a_{ij}^{p_i} \leqslant 1$.