Distances in PR, Audio and Image

Michel DEZA (ENS, Paris) and Elena DEZA (SPU, Moscow)

Séminaire Brillouin, IRCAM, 31 mai 2012

The abstraction of measurment, in terms of mathematical notions distance, similarity, metric, etc. was originated by M.Fréchet (1906) and F.Hausdorff (1914). Triangle inequality, crucial in it, appears already in Euclid.

Given a set X, a **distance** (or **dissimilarity**) on it is a function $d : X \times X \to \mathbb{R}_{\geq 0}$ with all $d(x,x) = 0$ and $d(x,y)=d(y,x)$ (**symmetry**).

A **similarity** is a symmetric function $s : X \times X \to \mathbb{R}_{\geq 0}$ such that $s(x,y) \leq s(x,x)$ holds for all $x, y \in X$ with equality if and only if $x = y$.

A **metric** is a symmetric function $d : X \times X \to \mathbb{R}_{\geq 0}$ with $d(x,y) = 0$ iff $x = y$ and **triangle inequality** $d(x,y) \leq d(x,z) + d(z,y)$ if $x, y, z \in X$.

A **metric space** is a set X with a metric defined on it: (X, d).

Main transforms used to obtain a **distance** d **from a similarity** $s \leq 1$ are: $d = \arccos s$, $d = -\ln s$, $d = 1 - s$, $d = \frac{1-s}{s}$, $d = \sqrt{1-s}$, $d = \sqrt{2(1-s^2)}$.

Metric spaces started century ago as a special case of an infinite topological space. But from K.Menger (1928) and L.M.Blumenthal (1953), an explosion of interest in both, finite and infinite metric spaces, occurred.

By now, theories involving distances and similarities florished in many areas of Mathematics including Geometry, Probability, Coding/Graph Theory. Many mathematical theories, in the process of their generalization, settled down on the level of metric space. It is ongoing process in Riemannian Geometry, Real Analysis, Approximation Theory.

On the other hand, distances and metrics are now an essential tool in many areas of Applied Mathematics, say, Clustering, Data Analysis, Statistics, Computer Graphics/Vision, Networks, Computational Biology.

Besides distances themselves, powerful distance-related notions and paradigms (various generalized metrics, metric transforms, numerical invariants, distance maps etc.) began to be applied.

CONTENTS

DISTANCES IN PATTERN RECOGNITION

1. PR SYSTEM, SUPERVISION, CLASSIFIERS
2. GENERALITIES ON PR DISTANCE FUNCTION
3. DISTANCES IN CLUSTERING
4. EXAMPLES OF PRACTICAL PATTERN RECOGNITION

PR (Pattern Recognition) is a part of Machine Learning (extraction of the regularity from the data) aiming to classify individual data (patterns, objects) into groups, categories, i.e., assosiate each data item with the label (say, "color") of a particular class.

For example, in optical character or handwriting PR (**spatial items**), image feature vectors (shape, gray shade etc.) are labeled by characters in the input, while in speech PR (**temporal items**), spectral or cepstral features of waveforms are labeled by phones or words.

PR system consists of sensor gathering and filtering raw data, (in form of measurements, observations), representatation computing information from data and classifier, a PR algorithm actually classifying, or describing data.

Representatation is feature-based if objects are given as points in $\mathbb{R}^n$ (feature vectors) once n features (parameters) are defined, or distance-based if objects are given by their distances (pairwise and to given ones) once a suitable distance function is defined.

- Supervised PR (instance-based, memory-based learning): there is a training set T of previously classified patterns (stored labeled examples, instances, templates, prototypes).
  Classifier algorithm generate a **classification** input-output **function**.
  PR system learn (approximate) behaviour of this function, which maps incoming pattern (query point, test point) into a class (the label of the best match) comparing it with given input-output examples from T.

  Main classifiers (decision rules) are: (relatively simple Bayesian, distance-based and decision boundary-based (decision trees, support vector machines, discriminant functions, powerful neural networks).

- Unsupervised PR learning: system itself establishes the classes, it does clustering, i.e. partitioning a data set into clusters defined by a suitable similarity (proximity) measure and so, by a distance function.

Main distance-based classifiers are: minimum distance (choose label of closest instance to query), k-nearest neighbor or k-NN (choose most frequently represented label among k nearest neighbors or decide voting weight by the distance of the neighbor), Parzen Windows (choose label with maximal occurence within a given size window (say, a ball) around query.

(Artificial) neural network is a Neuroscience-inspired computational model: the **neurons** (vertices, units, processing elements) are connected into an adaptive, learning complex network. A distance is important in two cases:

Radial Basis Function: output depends of the distance to a prototype and

Self-organizing Map where a set of neurons learn (unsupervised) to map input space I to less-dimensional output space preserving topology of I.

Another case of use of distances in PR: metric data structures indexing data in metric space; especially, metric trees used in Nearest Neighbor Search.

So, one of main PR challenges (for distance-based representation, classifier design, clustering etc.) is to define suitable distance function.
This choice is equivalent to implicit statistical assumption about data.

Compactness hypothesis (Arkadiev-Braverman, 1966): representations of similar objects should be close. So, their distance should be invariant with respect to small and irrelevant (preserving class) transformations of data.

Main unsolved PR problem is the relationship between data to be classified and the performance of PR classifiers. At present, design of such algorithms and, especially, distance functions is rather an art.

Inductive bias of any learning algorithm is the set of assumptions used to predict certain target outputs given some training examples demonstrating the intended input-output relation. For example, minimum description length: simpler hypotheses are more likely to be true (Occam's razor).

Cognitive bias: distortion of reality perception by humans (observer effects).

The distance between objects can be between their feature vectors (feature-based), or between their graphs (or other structural representations) or directly between raw data (as for image shapes).

Distance measure can be selected/interpreted in a **space**: Euclidean, other metric, non-metric including so general/exotic ones as kernels (positive-definite inner products), pseudo-Euclidean (indefinite inner product) and extended topology.

A distance d can be approximated by a metric: take $d(x, y)+d(y, x)$ if d is a quasi-metric ($d(x, y) \neq d(y, x)$) or eqv. classes if d is a semi-metric ($d(x, y)=0$ for $x \neq y$) and shortest paths if the triangle inequality fails.

A classifier then applied to obtained distance values,

either in this space (usually, neighborhood-based algorithms),

or in one where dimensions refer to distances to given objects,

or in less-dimensional space where original one is projected/embedded.

The **PR distance measures** are between models (statistical or prototypes) or between a model and observations.

For **Sound Recognition**, the distances are between a template reference and input signal, while for **Noise Reduction**, they are between original (reference) and distorted signal.

For **Image Retrieval**, the distances are between feature vectors of a query and reference, while for **Image Processing** (as Audio Noise Reduction), they are between approximated and "true" digital images (to evaluate algorithms).

Image Retrieval (similarity search) consists of (as for **pattern recognition** with other data: audio, DNA/protein sequences, text documents, time series etc.) finding images whose features values are similar either between them, or to given query or in given range.

Main ways of removing the irrelevant and redundant information in data :

Feature selection: most systems preselect a small (≤ 100, except automatic text PR) set of features based on intuition.

Dimensionality reduction projects high-dimensional data to a space of lower dimension (each new dimension is a combination of the original features) with minimal loss of information. Example: Principal Component Analysis

Space partitioning looks for **hyper-boxes** (regions of the sample or feature space, where the classes are represented with lower ambiguity. The best of these regions are used to create rules for a rule-based classier.

In feature extraction, new features are created by combinations and transformations of existing features.

Inductive decision trees are classiers recursively partitioning the space in order to create classication trees. For example, difference of entropy before and after the partition can be used to select the best feature.

Needed training set can be too large. For example, in handwritten digit PR, prototype digits of each class should come in all sizes, positions, angles, line thicknesses, writing styles, skews. So, given the set T of all **irrelevant** (i.e., preserving class) transformations, there are following solutions.

1. Feature extraction: find representation minimally affected by $t \in T$.

2. Design invariant distance d, i.e. $d(x, y) = d(t_1(x), t_2(y))$ for any pattern x, prototype y and $t_1, t_2 \in T$. Simard, Le Gun, Denker, Victorri (1998) tangent distance is, roughly, the distance between manifolds (or just curves) $T(x)$ and $T(y)$. (Procrustes distance between shapes is the case when T consists of translations and rotations of image.)

   But, in general, manifolds $T(x), T(y)$ have no analytic expression (hence, difficult to compute and store) and non-linear. So, the distance is taken between linear surfaces that best approximate them, tangents.

SVM (support vector machine) classifier (Vapnik, 1995) maps input vectors $x \in X \subset \mathbb{R}^n$ (with non-linear class boundaries) into elements $f(x) \in \mathbb{H}^N$ of a Hilbert space with $n \ll N \leq \infty$, so that boundaries became hyperplanes (linear) and inner product $\langle f(x), f(y) \rangle$ in $\mathbb{H}^N$ can still be computed in $\mathbb{R}^n$. It is possible if $K(x, y) := \langle f(x), f(y) \rangle$ is a **kernel** (symm. positive-definite function). Kernel trick: equiv. of linear PR in $\mathbb{H}^N$ to non-linear PR in $\mathbb{R}^n$.

For example, if $f(x = (x_1, x_2)) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, then $\langle f(x), f(y) \rangle = f(x)f(y)^T = (\langle x, y \rangle)^2 = K(x, y)$.

Main kernels used in SVM are: $x^T y$ (linear), $(ax^T y + b)^c$ (polynomial), $exp\{-a||x - y||^2\}$(radial basis function), $\tanh\{ax^y + b)\}$ (sigmoid) where $a > 0$ and b, c are positive integers.

Let $\{(x^{(i)}, y_i))\}_{i=1}^m$ be the set of **support vectors** $x^{(i)}$ (instances lying on 2 bounding hyperplanes: $\sum_{i=1}^m \lambda_i y_i = 0$, $\lambda_i > 0$) with labels $y_i \in \{-1, +1\}$ (two-class classification). Final SVM classifier $F(x)=t+\sum_{i=1}^m \lambda_i y_i K(x^{(i)}, x)$ (Cristianini and Shawe-Taylor, 2000) gives maximal margin of hyperplanes.

Instead of **fixing** distance/similarity function, it can itself be **learned**.

In Content-Based Image Retrieval, it is learned from a training set of positive and negative eqv. constraints ("similar" or "different" point pairs).

El-Naqa et al., 2004: this learning is seen as nonlinear regression of the similarity coefficient on the features of the image.

Hertz et al., 2005: such learning is by training binary classifiers (over pairs of images) with margins to distinguish between pairs from the same or different class. The signed margin is used as a distance function.

Eick et al., 2006: objects from data set are clustered by a given distance function D, then local class density information of each cluster is used by a weight adjustment heuristic to modify D so that density is increased in the attribute space. Process is repeated until "good" distance function is found.

Herz and Yanover, 2006: **peptid-peptid distance** was learned from pairs of peptids known to co-bind or not the same Human Leucocyte Antigen.

Main real-word PR applications are: Computer Vision (including Medical Imaging, Handwriting, Face Recognition), Biometric Identification, Audio (including Speech) Recognition, Biology, Credit scoring, Market Research, Geostatistics (including weather maps), Interner search engines, Text classification (spam/non spam, documents).

Especially intense ongoing search for suitable distances occurs in Image Analysis, Speech Recognition, Biology, Information Retrieval.

Clustering is applied in Computational Biology/Bioinformatics in order:

to build groups of genes with related expression pattern;

to group homologous sequences into gene families;

to automatically assign genotypes in high-troughput genotyping platforms;

and, in Ecology, to generate artificial phylogenies of organisms sharing some attributes at species/genus level or to make spatial/temporal comparisons of communities of organisms in heterogeneous environments.

Cluster Analysis consists of partition of data into relatively small number of **clusters**, i.e., such sets of objects, that (with respect of selected measure of distance) the objects, best possible "close" if belong to the same cluster, "far" if not, and further subdivision will impair above two conditions.

We give two typical examples of clustering. In **Information Retrieval** applications, nodes of peer-to-peer database network export a data (collection of text documents); each document is characterized by a vector from $\mathbb{R}^n$. An user **query** consists of a vector $x \in \mathbb{R}^n$, and user needs all documents in database which are **relevant** to it, i.e., belong to the ball in $\mathbb{R}^n$, centered in x, of fixed radius and with convenient distance function.

In **Record Linkage**, each document (database record) is represented by a term-frequency vector $x \in \mathbb{R}^n$ or a string, and one wants to measure semantic relevancy of syntactically different records.

Once a distance d between objects is selected, the **linkage metric**, i.e., a distance between clusters $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$ is usually one of the following:

average linkage: the average of the distances between the all members of those clusters, i.e., $\frac{\sum_i \sum_j d(a_i, b_j)}{mn}$;

single linkage (or **set-set distance**): the distance between the nearest members of those clusters, i.e., $\min_{i,j} d(a_i, b_j)$;

complete linkage: the distance between the furthest members of those clusters, i.e., $\max_{i,j} d(a_i, b_j)$;

centroid linkage: the distance between the **centroids** of those clusters, i.e, $||\tilde{a} - \tilde{b}||_2$, where $\tilde{a} = \frac{\sum_i a_i}{m}$, and $\tilde{b} = \frac{\sum_j b_j}{n}$;

Ward linkage: the distance $\sqrt{\frac{mn}{m+n}}||\tilde{a} - \tilde{b}||_2$.

A **data set** is a set of m n-sequences $(x_1^j, \ldots, x_n^j)$, $j = 1, \ldots, m$. The range $x_i^1, \ldots, x_i^m$ represent **attribute** S_i. It can be **numerical**, incl. **continuous** (real numbers) and **binary** (presence/absence expressed by 1/0), **ordinal** (numbers expressing rank only), or **nominal** (not ordered). Following setting of **distance-based machine learning** is used for many real-world applications with incomplete data and continuous+nominal attributes.

Given an $m \times (n+1)$ matrix $((x_{ij}))$, its row $(x_{i0}, x_{i1}, \ldots, x_{in})$ means **instance input vector** $x_i = (x_{i1}, \ldots, x_{in})$ with output class x_{i0}; the set of m instances represents a training set during learning. For any new input vector $y = (y_1, \ldots, y_n)$, the closest (in terms of selected distance d) instance x_i is sought, in order to **classify** y, i.e., predict its output class as x_{i0}. Then, say, $d(x_i, y) = \sqrt{\sum_{j=1}^n d_j^2(x_{ij}, y_j)}$ with $d_j(x_{ij}, y_j) = 1$ if x_{ij} or y_j is unknown. $d_j(x_{ij}, y_j) = 1_{x_{ij} \neq y_j}$ if **attribute** j (range of x_{ij}) is nominal; $d_j = |x_{ij} - y_j| / \max_t x_{tj} - \min_t x_{tj}$ if j continuous.

The choice of **similarities** and **distances** in Clustering depends on the nature of data and it is an art. Examples follow.

The **cosine similarity** (or **Orchini similarity**, **angular similarity**, **normalized dot product**) on $\mathbb{R}^n$ is

$$\frac{\langle x, y \rangle}{||x||_2 \cdot ||y||_2} = \cos \phi,$$

where ϕ is the angle between vectors x and y. In Record Linkage, it is is called **TF-IDF** (for term **Frequency** – **Inverse Document Frequency**).

The **cosine distance** is $1 - \cos \phi$.

The **Hamming metric** on $\mathbb{R}^n$ is $d_H = |\{i : 1 \leq i \leq n, x_i \neq y_i\}|$.

On vertices of unit cube $\{0,1\}^n$ it is l_1**-metric** and squared l_2**-metric**.
Eqv., for subsets $A, B \subset X$ with $|X| = n$, it is **measure metric** $|A \triangle B|$.

The **Bray-Curtis distance** on $\mathbb{R}^n$ is $\frac{\sum |x_i - y_i|}{\sum (x_i + y_i)}$.

The **Canberra distance** on $\mathbb{R}^n$ is $\sum \frac{|x_i - y_i|}{|x_i| + |y_i|}$.

The **Mahalonobis distance** (or **statistical distance**) on $\mathbb{R}^n$ is

$$\sqrt{(detA)^{\frac{1}{n}}(x-y)A^{-1}(x-y)^T},$$

where A is a positive-definite matrix.

The **Hellinger distance** on $\mathbb{R}^n_+$ is $\sqrt{2\sum\left(\sqrt{\frac{x_i}{\overline{x}}} - \sqrt{\frac{y_i}{\overline{y}}}\right)^2}$.

EXAMPLE 1.

In Face Recognition, are used sets of (vertical/horizontal) **cephalofacial dimensions**, i.e., distances between **fiducial** (used as a fixed standard of reference for measurement) facial points. The distances are normalized, say, with respect of **inter-pupillary distance** for horizontal ones.

For example, the following 5 independent facial dimensions are derived by Fellous, 1997, for facial gender recognition:

distance E between external eye corners,

nostril-to-nostril width N,

face wigth at cheek W

and two vertical distances: eye-to-eyebrow distance B and

distance L between eye midpoint and horizontal line of mouth.

In above terms, "femaleness" relies on large E, B and small N, W, L.

A typical example of real-world PR: Datcu and Rothkranz, 2007, proposed a Web based automatic emotion recognition system from audio/video data.

User can upload speech (in German) and visual (video sequence or photo) files and run remotely full emotion recognition process on the input face.

The output is one of 6 innate (to generate and interpret) facial expressions: happiness, anger, disgust, fear/anxiety, surprise (or boredom), sadness. All other expressions have to be learned by humans (Ekman and Friesen, 1978).

The visual data are encoded as vectors of 17 facial features. As above, they are selected 17 Euclidean distances between selected key 21 facial points.

The emotional content of speech data is evaluated using database Berlin of German emotional speech: utterance samples by 10 native German actors (5 females and 5 males) simulating emotions, were recorded at freq. 16kHz.

EXAMPLE 2.

In a computer, **processor** is the chip doing all the computations, and **memory** usually refers to **RAM** (random access memory). Processor **cache** stores small amounts of recently used information right next to the processor where it can be accessed much faster than memory.

The **reuse distance** (Mattson et al, 1970) of a memory location is the number of distinct memory references between two accesses of it. (Each one is counted only once because after access it is moved in cache.) This distance evaluate cache behavior of programs.

Cf. program **locality metric** (Gorla and Zhang, 1999) measuring globally the locations of program's components, their calls and the depth of nested calls by $\frac{\sum_{i,j} f_{ij} d_{ij}}{\sum_{i,j} f_{ij}}$, where d_{ij} is a distance between calling components i, j, and f_{ij} is the frequency of calls from i to j.

Reuse pattern is a histogram of the percentage of memory accesses whose reuse distance falls inside consecutive ranges $k\%$ divided between 0 and the **reuse data size** (maximal reuse distance). PR here (Ding and Zhong, 2003) detects whether the reuse pattern is predictable accross data inputs.

Reference histogram show the average reuse distance of each $k\%$ of all memory references. (Its use permits isolate effect of non-recurrent parts of the program and control the granularity of prediction.)

Given 2 reference histograms from 2 training data inputs, the formula for distance in i-th bin is $d_i = c_i + e_i f(s_i)$, where s_i is maximal reuse distance, f is known (at most linear, say, constant or linear) function. Coefficients c_i, d_i are computed from at least 2 training inputs (d_1, s_1), (d_2, s_2).

Limitations: predicting reuse pattern does not mean predicting execution time; the prediction gives the percentage distribution but not the total number of memory accesses.

EXAMPLE 3: **distance function selection** for PR in neuronal network.

To gain information about functional connectivity of a neuronal network, one needs to classify neurons, in terms of their firing similarity; so, to select a distance function and a clustering algorithm. A classical example: simple and complex cells discrimination between in the primary visual cortex.

A human brain has $\approx 10^{11}$ of **neurons** (nerve cells). Neuronal response to a stimulus is a continuous time series. It can be reduced, by a threshold criterion, to much simpler discrete series of **spikes** (short electrical pulses),

A **spike train** is a sequence $x = (t_1, \ldots, t_s)$ of s events (neuronal spikes, or hearth beats, etc.) listing absolute spike times or inter-spike time intervals.

"Good" **distances between spike trains** should minimize bias (due to predefining analysis parameters if any) and resulting clusters should well match the stimuli and reproduce some control clustering.

Main **distances between spike trains** $x = x_1, \ldots, x_m$ and $y = y_1, \ldots, y_n$:

1. $\frac{|n-m|}{\max\{m,n\}}$ (**spike count distance**); no bias by predefining analysis parameters, but the temporal structure of trains is missed.

2. $\sum_{1 \le i \le s}(x'_i - y'_i)^2$, where, say, $x' = x'_1, \ldots, x'_s$ is the sequence of local firing rates of train $x = x_1, \ldots, x_m$ partitioned in s time intervals of length T_{rate} (**firing rate distance**); bias due to predefinition of T_{rate}.

3. Let $\tau_{ij} = \frac{1}{2}\min\{x_{i+1} - x_i, x_i - x_{i-1}, y_{i+1} - y_i, y_i - y_{i-1}\}$ and $c(x|y) = \sum_{i=1}^{m}\sum_{j=1}^{n} J_{ij}$, where $J_{ij} = 1, \frac{1}{2}, 0$ if $0 < x_i - y_i \le \tau_{ij}$, $x_i = y_i$, else, resp. **Event sinchronization distance** (Quiroga et al., 2002) is $1 - \frac{c(x|y)+c(y|x)}{\sqrt{mn}}$. Two metrics (above and below) have no parameter presetting time scale.

4. Let $x_{isi}(t) = \min\{x_i : x_i > t\} - \max\{x_i : x_i < t\}$ for $x_1 < t < x_m$, and let $I(t) = \frac{x_{isi}(t)}{y_{isi}(t)-1}$ if $x_{isi}(t) \le x_{isi}(t)$ and $I(t) = 1 - \frac{y_{isi}(t)}{x_{isi}(t)}$, otherwise. Kreuz et al., 2007, **ISI distances** are $\int_{t=0}^{T} dt|I(t)|$ and $\sum_{i=1}^{m} |I(t_i)|$.

5. **information distances** (**Kullback-Leibler distance** $\sum_i x_i \ln \frac{x_i}{y_i}$ or **Kolmogorov complexity** $K(x|y)$ of train x given train y, i.e., the length of the shortest program to compute x if y is provided as an auxiliary input.

The **Kolmogorov complexity** (**algotithmic entropy**) $K(x)$ of x is the length of a shortest program x^* (ultimate compressed version of x) to compute x on an universal computer usung a **Turing-complete** language.

6. The **Lempel-Ziv distance** between two binary n-strings x and y is $\max\{\frac{LZ(x|y)}{LZ(x)}, \frac{LZ(y|x)}{LZ(y)}\}$, where $LZ(x) = \frac{|P(x)| \log |P(x)|}{n}$ approximates uncomputable **Kolmogorov complexity** $K(x)$, and $LZ(x|y) = \frac{|P(x) \backslash P(y)| \log |P(x) \backslash P(y)|}{n}$. Here $P(x)$ is the set of non-overlapping substrings into which x is parsed sequentially, so that new substring is not yet contained in the set of substrings generated so far. For example, such **Lempel-Ziv parsing** for $x = 001100101010011$ is $0|01|1|00|10|101|001|11$.

7. **Victor-Purpura distance** is the minimal cost of transforming x into y by operations: insert, delete, shift a spike by time t with costs $1, 1, qt$.

8. **van Rossum distance**, 2001, is $\sqrt{\int_0^\infty (f_t(x) - f_t(y))^2) dt}$, where x is convoluted with $h_t = \frac{1}{\tau} e^{-t/\tau}$ and $\tau \approx 12$ ms (best); $f_t(x) = \sum_0^m h(t - x_i)$.

Victor-Purpura distance $\approx$ van Rossum L_1-distance with $h_t = \frac{q}{2}$ if $0 \leq t < \frac{2}{q}$

9. **Aronov et al. distance** between two sets of labelled (by firing neuron) spike trains is the minimal cost of transforming one to the other by spike operations insert/delete, shift by time t, relabel with costs 1, qt, k, resp.

PR of **3D structures**, besides Image Analysis and Tomography, applied mainly in following areas:

In Biology (from 1982), to predict protein **secondary structure** (roughly, the set of helices, or the list of paired bases, making up protein) and **tertiary structure** (geometric form protein takes in space) from multiple aligned **primary structures** (amino acid sequences).

In **pharmacore** (minimum active sequence) identification and drug design (from 1986).

In chemical reactivity studies (from 1987).

For 3D molecular template recognition (from 1991): molecular shape similarities (from interatomic distances) and molecular electronic similarities (comparaison of their density functions).

General observations on **distance design** follow.

- Distance should be **invariant** with respect to small and irrelevant transformations of data.
- Distance can be **upgraded** to metric and **corrected** for bias.
- Distance can be between a **prototype** and **input** (or query) or between **true** and **distorted** (or approximated) data.
- Distance can be **abstract** $\mathbb{R}_{\geq 0}$-valued or **physical** (as length): between **spatial** or **temporal** points (the length of journey till, say, speed or concentration reach fixed value).
- Usually, **several** distances on the same data should be compared.
- Distance/similarity can be **implicite** as in Clustering.
- Instead of **fixing** distance/similarity function, it can be **learned**.
- The choice of good distance/similarity is rather an art.

BIRDVIEW ON METRIC SPACES

1. Metric repairs
2. Generalizations of metric spaces
3. Transform metrics
4. Numeric invariants of metric spaces
5. Relevant notions: special subsets, mappings, curves, convexity
6. Main classes of metric spaces

METRIC REPAIRS

Let X be a set. A function $d : X \times X \to \mathbb{R}_{\geq 0}$ with all $d(x, x) = 0$ is called a **quasi-distance** (or, in

Topology, **premetric**) on X.

A quasi-distance d is a **quasi-semi-metric** (or **hemimetric**) if it holds $d(x, y) \leq d(x, z) + d(z, y)$ (**triangle inequality**) for all $x, y, z \in X$, and

it is a **quasi-metric** if, moreover, $d(x, y) > 0$ for $x \neq y$.

If $d(x, y) = d(y, x)$ (**symmetry**) holds for all $x, y \in X$, then above quasi-distance, quasi-semi-metric, quasi-metric become, respectively: **distance** (or **dissimilarity**), **semi-metric** (or **pseudo-metric**), **metric**.

For a **distance** d, the function, defined by $D(x,x)=0$ and, for $x \neq y$ by $D_1(x,y) = d(x,y) + \max_{x,y,z \in X}(d(x,y) - d(x,z) - d(y,z))$ is a semi-metric. Also, $D_2(x,y) = d(x,y)^c$ is a semi-metric for sufficiently small $c \geq 0$. Also, the function $D_3(x,y) = \inf \sum_i d(z_i, z_{i+1})$, where the infimum is taken over all sequences $x = z_0, \ldots, z_{n+1} = y$, is the shortest path semi-metric.

For a **semi-metric** d on X, define equivalence relation by $x \sim y$ if $d(x,y) = 0$; let $[x]$ be the equivalence class containing x. Then $D([x],[y]) = d(x,y)$ is a metric on the set $\{[x] : x \in X\}$ of classes.

For a **quasi-metric** d, functions $\max\{d(x,y), d(y,x)\}$, $\min\{d(x,y), d(y,x)\}$ and $\frac{(d^p(x,y)+d^p(y,x))^{\frac{1}{p}}}{2}$ (usually, $p=1$) are symmetrization metrics; they are equivalent, i.e., define the same topology.

For a **metric** d, the function $D(x,y) = \frac{d(x,y)}{1+d(x,y)} < 1$, is a 1-bounded metric.

1. $d(x,y) \leq d(x,z) + d(z,y)$ (**triangle inequality**), i.e., a **metric**;

2. $d(x,y)d(u,z) \leq d(x,u)d(y,z) + d(x,z)d(y,u)$, a **Ptolemaic metric**;

3. $d(x,y) + d(z,u) \leq \max(d(x,z) + d(y,u), d(x,u) + d(y,z))$ (**4-point inequality**), a $\mathbb{R}_{>0}$**-edge-weighted tree metric** (it is 2, 5, 7);

4. $d(x,y) \leq \max(d(x,z), d(z,y))$, an **ultrametric** (it is 3);

5. $d(x,y) + d(z,u) \leq 2\delta + \max\{d(x,z) + d(y,u), d(x,u) + d(y,z)\}$ **for** $\delta \geq 0$, a δ**-hyperbolic metric**;

6. $d(x,y) \leq d(x,z) + d(z,y) - d(x,z)d(z,y)$ (**equivalent to** $1 - d(x,y) \geq (1 - d(x,z))(1 - d(z,y))$), a P**-metric**;

7. $\sum_{1 \leq i < j \leq n} b_i b_j d(x_i, x_j) \leq 0$ **for** $b \in \mathbb{Z}^n$, $\sum_{i=1}^n b_i = 1$, **a hypermetric**;

8. $d(x,y) \leq C(d(x,z) + d(z,y))$ **for a constant** $C \geq 1$, a **near-metric**;

9. $d(x,y) \leq d(x,z) + d(z,y) - d(z,z)$ **for** $0 \leq d(z,z) \leq \inf_u d(z,u)$**, i.e., self-distances are small**, a **partial metric**.

OTHERS GENERALIZATIONS OF METRIC SPACES

A **topological space** (X, τ) is a set X with a **topology** τ, i.e., a collection of subsets of X (called **open sets**), with the following properties:

1. $X \in \tau$, $\emptyset \in \tau$;
2. If $A, B \in \tau$, then $A \cap B \in \tau$;
3. For any collection $\{A_\alpha\}_\alpha$, if all $A_\alpha \in \tau$, then $\cup_\alpha A_\alpha \in \tau$.

Any metric space (X, d) generates a **metric topology** consisting of all **open balls** $B(x, r) = \{y \in X : d(x, y) < r\}$.

Two metrics d_1 and d_2 on a set X are called **equivalent** if they define the same topology on X, i.e., if, for every $x_0 \in X$, every open metric ball with center at x_0 defined with respect to d_1, contains an open metric ball with the same center but defined with respect to d_2, and conversely. All metrics on a finite set are equivalent; they generate the **discrete topology**.

A **resemblance** is a symmetric function $d : X \times X \to \mathbb{R}$ such that: either all $d(x, x) \leq d(x, y)$ holds (then d is called **forward resemblance**), or all $d(x, x) \geq d(x, y)$ holds (then d is called **backward resemblance**).

Any resemblance d induces a **strict partial order** $\prec$ on unordered pairs of elements of X by defining $\{x, y\} \prec \{u, v\}$ iff $d(x, y) < d(u, v)$. For backward resemblance d, the forward one $-d$ induces the same partial order.

A **2-metric** is function $d : X \times X \times X \to \mathbb{R}_{\geq 0}$ which is **totally symmetric** (i.e., $d(x_1, x_2, x_3)$ is unchanged by any permutation of arguments), **zero conditioned** (i.e., $d(x_1, x_2, x_3) = 0$ iff $x_i = x_j$ for some $1 \leq i < j \leq 3$) and satisfy **tetrahedron inequality**

$$d(x_1, x_2, x_3) \leq d(x_4, x_2, x_3) + d(x_1, x_4, x_3) + d(x_1, x_2, x_4).$$

A **m-metric** (or **m-volume**) is defined by **m-simplex inequality**. The cases $m = 1, 2$ correspond to usual metric (length) and area, respectively.

The **pseudo-Euclidean distance** of signature $(p, q = n - p)$ on $\mathbb{R}^n$ is

$$d_{pE}(x, y) = \sum_{i=1}^{p} (x_i - y_i)^2 - \sum_{i=p+1}^{n} (x_i - y_i)^2.$$

The pseudo-Euclidean space of signature $(p, q = n - p)$ is a real vector space equipped with a non-degenerate, indefinite, symmetric bilinear function $\langle \cdot, \cdot \rangle$. A basis $e_1, \ldots, e_{p+q}$ is orthonormal if $\langle e_i, e_j \rangle = 0$ for $i \neq j$, $\langle e_i, e_i \rangle = +1$ for $1 \leq i \leq p$ and $\langle e_i, e_i \rangle = -1$ for $p + 1 \leq i \leq p + q$.

Given an orthonormal basis, the inner product of two vectors x and y is $\langle x, y \rangle = \sum_{i=1}^{p} x_i y_i - \sum_{i=p+1}^{p+q} x_i y_i$.

The pseudo-Euclidean space can be seen as $\mathbb{R}^p \times i\mathbb{R}^q$, where $i = \sqrt{-1}$.

The "norm" $\langle x, x \rangle$ of non-zero vector x can be positive, negative or zero; then x is called **space**, **time** or **light** vector, respectively.

The case $(p, q) = (1, 3)$ is used as space-time model of Special Relativity.

An **uniform space** (Weil, 1937) is a set X with a non-empty collection $\mathcal{U}$ of subsets of $X \times X$ (**entourages**) such that it holds:

1. Every subset of $X \times X$ which contains a set of $\mathcal{U}$, belongs to $\mathcal{U}$;
2. Every finite intersection of sets of $\mathcal{U}$ belongs to $\mathcal{U}$;
3. Every set $V \in \mathcal{U}$ contains the set $\{(x, x) : x \in X\} \subset X \times X$ (**diagonal**);
4. If V belongs to $\mathcal{U}$, then the set $\{(y, x) : (x, y) \in V\}$ belongs to $\mathcal{U}$;
5. If V belongs to $\mathcal{U}$, then there exists $V^{'} \in \mathcal{U}$ such that $(x, z) \in V$, whenever $(x, y), (y, z) \in V^{'}$.

Every **metric space** (X, d) is uniform: an entourage in (X, d) is a subset of $X \times X$ containing $V_\epsilon = \{(x, y) \in X \times X : d(x, y) < \epsilon\}$ for some $\epsilon > 0$.

Other basic example of uniform space are **topological groups**.

Every uniform space $(X, \mathcal{U})$ generate a topology: all sets $A \subset X$ such that, for any $x \in A$, there is a set $V \in \mathcal{U}$ with $\{y : (x, y) \in V\} \subset A$.

Every uniformity induces a **proximity** σ where $A \sigma B$ if and only if $A \times B$ has non-empty intersection with any entourage.

A **proximity space** is a set X with a **proximity**, i.e., symmetric binary relation δ on the **power set** $P(X)$ (of all its subsets) with $A \delta A$ iff $A \neq \emptyset$ and $A \delta (B \cup C)$ if and only if $A \delta B$ or $A \delta C$ (**additivity**).

Every **metric space** (X, d) is a proximity space: define $A \delta B$ iff $d(A, B) = \inf_{x \in A, y \in B} d(x, y) = 0$.

An **approach space** (Lowe, 1989) is a pair (X, D), where X is a set, and D is a **point-set distance**, i.e., a function $D(x, A) \geq 0$ of $x \in X$ and $A \subset X$ satisfying, for all $x \in X$ and all $A, B \subset X$, to:

1. $D(x, \{x\}) = 0$;
2. $D(x, \{\emptyset\}) = \infty$;
3. $D(x, A \cup B) = \min\{D(x, A), D(x, B)\}$;
4. $D(x, A) \leq D(x, A^\epsilon) + \epsilon$, for any $\epsilon \geq 0$
   (here $A^\epsilon = \{x : D(x, A) \leq \epsilon\}$ is "ϵ-ball" with the center x).

Any **metric space** (X, d) (moreover, any quasi-semi-metric space) is an approach space with $D(x, A) = \min_{y \in A} d(x, y)$ (the usual point-set distance).

Consider a set X and a map $cl : P(X) \to P(X)$ with $cl(\emptyset) = \emptyset$. The maps $cl(A)$ (for $A \subset X$), its dual $int(A) = X \backslash cl(X \backslash A)$ and $N : X \to P(X)$ with $N(x) = \{A \subset X : x \in int(A)\}$ are called **closure**, **interior** and **neighborhood** map, resp. A subset $A \subset X$ is **closed** if $A = cl(A)$ and **open** if $A = int(A)$. Consider the following possible properties of (X, cl):

1. $A \subseteq B$ implies $cl(A) \subseteq cl(B)$ (**isotony**);
2. $A \subseteq cl(A)$(**enlarging**);
3. $cl(A \cup B) = cl(A) \cup cl(B)$ (**linearity**, and, in fact, 3. implies 1.);
4. $cl(cl(A)) = cl(A)$ (**idempotency**).

The pair (X, cl) is called **extended topology** if 1. hold, **Brissaud space** (Brissaud, 1974) if 2. hold, **neighborhood space** (Hammer, 1964) if 1., 2. hold, **Smyth space** (Smyth, 1995) if 3. hold, **pretopology** (Čech, 1966) if 2., 3. hold, and **closure space** (Soltan, 1984) if 1., 2, 4. hold.

(X, cl) is usual topology, in closure terms, if 2., 3., 4. hold.

METRIC TRANSFORMS

A **transform metric** is a metric on a set X which is obtained as a function of a given metric (or metrics) on X. Examples obtained from a given metric d (or metrics d_1 and d_2) on X follow (here $t > 0$):

1. $td(x, y)$ (t-**scaled metric**, or **dilated metric**);
2. $\min\{t, d(x, y)\}$ (t-**truncated metric**, or **t-bounded metric**);
3. $\max\{t, d(x, y)\}$ for $x \neq y$ (t-**discrete metric**);
4. $d(x, y) + t$ for $x \neq y$ (t-**translated metric**);
5. $\frac{d(x,y)}{1+d(x,y)}$;
6. $\max\{d_1(x, y), d_2(x, y)\}$;
7. $\alpha d_1(x, y) + \beta d_2(x, y)$, where $\alpha, \beta > 0$ (so, **semi-metric cone** on X);
8. $d^z(x, y) = \frac{d(x,y)}{d(x,z)+d(y,z)+d(x,y)}$ where z is an fixed element of X (**biotope transform metric**).

- Given a metric space (X, d) and $0 < \alpha \leq 1$, the **power transform metric** (or **snowflake transform metric**) on X is $(d(x, y))^{\alpha}$.

  It is a **metric**, for any positive α if and only if d is an **ultrametric**.

- Given a metric space (X, d) and a point $z \in X$, the **involution transform metric** on $X \backslash \{z\}$ is

$$d_z(x, y) = \frac{d(x, y)}{d(x, z)d(y, z)}.$$

  It is a **metric**, for any $z \in X$, if and only if d is a **Ptolemaic metric**.

- Given a metric space (X, d) and $\lambda > 0$, the **Schoenberg transform metric** on X is

$$D(x, y) = 1 - e^{-\lambda d(x,y)},$$

  $D(x, y$ are *P*-**metrics**, i.e. $D(x, y) \leq D(x, z)+D(z, y)$-$D(x, z)D(z, y)$.

- An **induced metric** is a restriction of a metric (X, d) to $X' \subset X$.
- Given metric spaces (X, d_X), (Y, d_Y) and injective mapping $g : X \to Y$, the **pullback metric** (of (Y, d_y) by g) on X is $d_Y(g(x), g(y))$.
- Given a metric space (X, d) and an equivalence relation $\sim$ on X, the **quotient semi-metric** on the set $\overline{X} = X/\sim$ of equivalence classes is $\overline{d}(\overline{x}, \overline{y})$=$\inf_{m\in\mathbb{N}} \sum_{i=1}^{m} d(x_i, y_i)$, where the infimum is over all sequences $x_1, y_1, \ldots, x_m, y_m$ with $x_1 \in \overline{x}$, $y_m \in \overline{y}$ and $y_i \sim x_{i+1}$ if $1 \leq i =\leq m-1$
- Given $n \leq \infty$ metric spaces (X_1, d_1), (X_2, d_2), $\ldots$, (X_n, d_n), the **product metric** is any metric on their **Cartesian product** $X_1 \times X_2 \times \cdots \times X_n = \{x = (x_1, x_2, \ldots, x_n) : x_1 \in X_1, \ldots, x_n \in X_n\}$, defined as a function of $d_1, \ldots, d_n$.

- Given a metric space (X, d) and a point $z \in X$, the **Farris transform metric** on $X \backslash \{z\}$ is defined by $D_z(x, x) = 0$ and, for $x \neq y$, by

$$D_z(x, y) = C - (x.y)_z,$$

where $C > 0$ is a constant and $(x.y)_z = \frac{1}{2}(d(x, z) + d(y, z) - d(x, y))$ is the **Gromov product**. It is a **metric** if and only if $C \geq C_0$ for some number $C_0 \in (\max_{x,y \in X \backslash \{z\}, x \neq y}(x.y)_z, \max_{x \in X \backslash \{z\}} d(x, z)]$.

Farris transform is an **ultrametric** if and only if d is a $\mathbb{R}_{>0}$**-edge-weighted tree metric**.

In Phylogenetics, where it was applied first, the term *Farris transform* is used for function $d(x, y) - d(x, z) - d(y, z)$.

- Given a metric space (X, d) with any points $x, y \in X$ joined by a **rectifiable curve** (i.e., of finite length), the **intrinsic metric** $D(x, y)$ is the infimum of the lengths of rectifiable curves connecting x and y.

  A (metric) **curve** γ is a continuous mapping $\gamma : I \to X$ from an interval I of $\mathbb{R}$ into X. The **length** $l(\gamma)$ of a curve $\gamma : [a, b] \to X$ is

  $$l(\gamma) = \sup\{ \sum_{1 \le i \le n} d(\gamma(t_i), \gamma(t_{i-1})) : n \in \mathbb{N}, a = t_0 < t_1 < \cdots < t_n = b\}).$$

- The **Riemannian metric** of a connected n-dim. smooth **manifold** M^n, is a collection of positive-definite symmetric bilinear forms $((g_{ij}))$ on the tangent spaces of M^n which varies smoothly from point to point.

  The length of a curve γ on M^n is $\int_\gamma \sqrt{\sum_{i,j} g_{ij} dx_i dx_j}$.

  The **Riemannian distance** (between two points of M^n) is intrinsic metric on M^n, i.e. the infimum of lengths of curves, connecting them.

NUMERICAL INVARIANTS OF METRIC SPACES

- For any $p, q > 0$, let $M_p^q(X) = \inf \sum_{i=1}^{+\infty} (diam A_i)^p$, where infimum is taken over all countable coverings $\{A_i\}$ of X with diameter of $A_i < q$.

  The **Hausdorff dimension** (or **fractal dimension**) of X is

  $$\mathbf{dim_{Haus}} = \inf\{p : \lim_{q \to 0} M_p^q(X) = 0\}$$

  .

- For any compact metric space (X, d), its **topological dimension** is

  $$\mathbf{dim_{top}(X, d)} = \inf_{d'}(\dim_{Haus}(X, d')),$$

  where d' is any metric on X topologically equivalent to d.

  Two metrics d_1, d_2 on a set X are **equivalent** if they define same **topology** on X (for any $x_0 \in X$, any open d_1- metric ball centered at x_0 contains an open d_2-metric ball centered at x_0 and conversely).

- A **fractal** is a metric space for which $dim_{top} < dim_{Haus}$.

- For a metric space (X, d) and any $q > 0$, let $N_X(q)$ be the minimal number of sets with diameter $\leq q$ needed in order to cover X. The number $\mathbf{dim_{metr}} = \lim_{q \to 0} \frac{\ln N(q)}{\ln(1/q)}$ (if it exists) is called its **metric dimension** (or **Minkowski dimension**, **box-counting dimension**).

  For a totally bounded (X, d), it holds $dim_{top} \leq dim_{Haus} \leq dim_{metr}$. Any $X \subset \mathbb{E}^n$ with $Int\ X \neq \emptyset$ has $dim_{Haus} = dim_{metr}$.

- The **Assouad-Nagata dimension** $\mathbf{dim_{AN}}$ of a metric space (X, d) is the smallest integer n for which there exist a constant $C > 0$ such that, for all $s > 0$, there exists a covering of X by its subsets of diameter at most Cs with no point of X belonging to more than $n + 1$ elements.

  d called a **doubling metric** if $dim_{AN} < \infty$. It holds $dim_{top} \leq dim_{AN}$.

RADII S OF METRIC SPACE

If (X,d) is A-**bounded** ($A = \sup_{x,y\in X} d(x,y) < \infty$), then A is its **diameter**. If (X,d) is a-**discrete** ($a = \inf_{x,y\in X, x\neq y} d(x,y) > 0$), then $\frac{A}{a}$ is its **metric spread** (or **aspect ratio**, **normalized diameter**).

- The **metric radius** of (X,d) is $r = \inf_{x\in X} \sup_{y\in X} d(x,y)$. It holds $\frac{A}{2} \leq r \leq A$; some authors call *radius* the half-diameter.
- Given a subset $M \subseteq X$ of bounded (X,d),

  its **metric hull** the intersection of all closed metric balls containing M,

  its **covering radius** (or **directed Hausdorff distance**) is $d_{dHaus}(X,M) = \sup_{x\in X} \inf_{y\in M} d(x,y)$,

  its **Chebyshev radius** (or **remoteness**) is $\inf_{x\in X} \sup_{y\in M} d(x,y)$,

  its **packing radius** is $\sup\{r : \inf_{x,y\in M, x\neq y} d(x,y) > 2r\}$.

- A metric space (X, d) has the **order of congruence** n if every finite metric space which is not **isometrically embeddable** in (X, d) has a subspace with $\leq n$ points which is not isometrically embeddable in it.
- Given a compact connected metric space (X, d), there exists a unique **rendez-vous number** $r(X, d) > 0$, such that for all $x_1, \ldots, x_n \in X$ and any n, there exists an $x \in X$ with $\frac{1}{n} \sum_{i=1}^{n} d(x_i, x) = r(X, d)$.
- Given a set $D \subset \mathbb{R}_{>0}$, the **$D$-chromatic number** of (X, d) is the standard **chromatic number** of the **D-distance graph** of (X, d), i.e., the graph with the vertex-set X and the edge-set $\{xy : d(x, y) \in D\}$.
- The **average distance** is the number $\frac{1}{|X|(|X|-1)} \sum_{x,y \in X} d(x, y)$.
  The **Wiener index** (used in Chemistry) is $\frac{1}{2} \sum_{x,y \in X} d(x, y)$.
- For $s \neq 0$, the **s-energy** (or **unnormalized $\frac{1}{s}$-moment**) is $\sum_{x,y \in X, x \neq y} \frac{1}{d^p(x,y)}$. **0-energy** is $-\log \prod_{x,y \in X, x \neq y} d(x, y)$.

Given a metric space (X, d) and $s > 0$, the **Frechét function** is $F_s(x) = \mathbb{E}[d^s(x, y)]$. For a finite $M \subset X$, $F_s(x) = \sum_{y \in M} w(y) d^s(x, y)$. , where $w(y)$ is a weight function. The points minimizing $F_1(x)$ and $F_2(x)$ are called **Frechét median** and **Frechét-Karcher mean**.

For $(X, d) = (\mathbb{R}^n, ||x - y||_2)$ and equal weights, these points are called the **geometric median** (or **Fermat-Weber point**, **1-median**) and the **geometric center** (or **centroid**, **barycenter**, **center of mass**).

For $(X, d) = (\mathbb{R}_{>0}, |f(x) - f(y)|)$, where $f : \mathbb{R}_{>0} \to \mathbb{R}$ is injective and continuous, the Frechét mean of $M \subset \mathbb{R}_{>0}$ is the **Kolmogorov mean** (or f**-mean**) $f^{-1}(\frac{\sum_{x \in M} f(x)}{|M|})$. It is **arithmetic**, **geometric**, **harmonic** and **power mean** if $f = x, \log(x), \frac{1}{x}$ and $f = x^p$ $(p \neq 0)$ which is quadratic, arithmetic, geometric, harmonic mean and maximum, minimum for $p = 2, = 1, \to 0, \to -1$, and $\to +\infty, \to -\infty$.

RELEVANT NOTIONS: SUBSETS, MAPPINGS, CURVES, CONVEXITY

- Given distinct points $x, y \in X$, the **midset** (or **bisector**) is the set $\{z \in X : d(x,z) = d(y,z)\}$ of **midpoints** z.
- $M \subset X$ is a **metric basis** of X if $d(x,z) = d(y,z)$ for all $z \in M$ implies $x = y$. The numbers $d(x,z), z \in M$, are the **metric coordinates** of x.
- Given a finite or countable semi-metric space $(X = \{x_1, \cdots, x_n\}, d)$, its **distance matrix** is the symmetric $n \times n$ matrix $((d_{ij}))$, where $d_{ij} = d(x_i, x_j)$ for any $1 \leq i, j \leq n$.

  The **semi-metric cone** is the set of all distance matrices on X.
- The **proximity** (or **underlying**) **graph of** metric space (X, d) is a graph with the vertex-set X and xy being an edge if no point $z \in X$ with $d(x,y) = d(x,z) + d(z,y)$ exists.

- The **point-set distance** $d(x, M)$ between $x \in X$ and $M \subset X$ is $\inf_{y \in M} d(x, y)$. The function $f_M(x) = d(x, M)$ is **distance map**. Distance maps are used in MRI (M being gray/white matter interface) as cortical maps, in Image Processing (M being image boundary), in Robot Motion (M being the set of obstacle points).

- A subset $M \subset X$ is **Chebyshev set** if for every $x \in X$, there is **unique** $z \in M$ with $d(x, z) = d(x, M)$.

- The **set-set distance** between two subsets $A, B \subset X$ is $\inf_{x \in A,} d(x, B) = \inf_{x \in A, y \in B} d(x, y)$. In Cluster Analysis, it is **single linkage**, while $\sup_{x \in A, y \in B} d(x, y)$ is **complete linkage**.

- The **Hausdorff metric** (on all compact subspaces of (X, d)) is $d_{Haus}(A, B) = \max\{d_{dHaus}(A, B), d_{dHaus}(B, A)\}$ where $d_{dHaus}(A, B)$ is $\max_{x \in A} \min_{y \in B} d(x, y)$, i.e., the **directed Hausdorff distance**.

MAPPINGS FOR METRIC SPACES

- Given metric spaces (X, d_X) and (Y, d_Y), a function $f : X \to Y$ is an **isometric embedding** of X into Y if it is injective and $d_Y(f(x), f(y)) = d_X(x, y)$ holds for all $x, y \in X$.

  An **isometry** is a bijective isometric embedding.

- Two metric spaces (X, d_X) and (Y, d_Y) are **homeomorphic** if there exists a bijection $f : X \to Y$ with **continuous** f and f^{-1} , i.e., all points close to x map to points close to $g(x)$.

- Given metric spaces (X, d_X) and (Y, d_Y), a function $f : X \to Y$ is called a **short mapping** from X to Y if, for all $x, y \in X$, holds $d_Y(f(x), f(y)) \leq d_X(x, y)$. The **category of metric spaces** (Isbell), denoted by Met, has metric spaces as objects and **short mappings** as morphisms. In Met, the **isomorphisms** are **isometries**.

- Again, given metric spaces (X, d_X) and (Y, d_Y), a function $f : X \to Y$ is an **isometric embedding** of X into Y if it is injective and $d_Y(f(x), f(y)) = d_X(x, y)$ holds for all $x, y \in X$.

  An **isometry** is a bijective isometric embedding.

- A function $f : X \to Y$ is a **quasi-isometry** if there are numbers $C > 1$ and $c > 0$ such that $C^{-1}d_X(x, y) - c \leq d_Y(f(x), f(y)) \leq Cd(x, y) + c$, and for every point $y \in Y$ there is a point $x \in X$ with $d_Y(y, f(x)) \leq c$.

  A quasi-isometry with $C = 1$ is **coarse** (or **rough**) **isometry**.

- A metric space (X, d) is **homogeneous** if, for each two finite isometric subsets $Y = \{y_1, \ldots, y_m\}$ and $Z = \{z_1, \ldots, z_m\}$ of X, there exists a self-isometry (motion) of (X, d) mapping Y to Z.

- (X, d) is **symmetric** if for any $p \in X$ there is a **symmetry relative to** p, i.e., a **motion** (self-isometry) f_p of (X, d) such that $f_p(f_p(x)) = x$ for all $x \in X$ and p is an isolated fixed point of f_p.

CURVES AND CONVEXITY

Given a metric space (X, d), a **curve** γ is a continuous map $\gamma : I \to X$ from an interval $I \subset \mathbb{R}$. The **length** of a curve $\gamma : [a, b] \to X$ is

$$l(\gamma) = \sup\{ \sum_{1 \le i \le n} d(\gamma(t_i), \gamma(t_{i-1})) : n \in \mathbb{N}, a = t_0 < t_1 < \cdots < t_n = b\}.$$

- The **internal metric** of (X, d) is $d_i(x, y) = \inf l(\gamma)$ over all curves $\gamma(x, y) : [0, 1] \to X$ with $\gamma(0) = x$, $\gamma(1) = y$ and $l(\gamma) < \infty$. If $d = d_i$, then d is called **intrinsic metric** and (X, d) **length space**.

- If, moreover, any two points x, y are joined by a **shortest path** (an isometric embedding $\gamma : [0, 1] \to X$ with $\gamma(0) = x$, $\gamma(1) = y$), then d is called **strictly intrinsic** and (X, d) **geodesic space**.

  The curve $\gamma(x, y)$ is called a **geodesic** (or **locally shortest**, **locally isometric**) if $l(\gamma(x, y)) = d(x, y)$.

- A complete metric space (X, d) is geodesic iff it is **midpoint convex**, i.e., for any $x, y \in X$, $x \neq y$, there is a **midpoint** $z = m(x, y) \in X$ with $d(x, y) = d(x, z) + d(z, y)$ and $d(x, z) = \frac{1}{2} d(x, y)$.

- Midpoint convex (X, d) is **Busemann convex** if for any $x, y, z \in X$ and midpoints $m(x, z)$ and $m(y, z)$, it holds

$$d(m(x, z), m(y, z)) \leq \frac{1}{2} d(x, y).$$

- Midpoint convex (X, d) is **ball convex** if for all $x, y, z \in X$ it holds

$$d(m(x, y), z) \leq \max\{d(x, z), d(y, z)\}.$$

- Midpoint convex (X, d) is **distance convex** if for all $x, y, z \in X$ holds

$$d(m(x, y), z) \leq \frac{1}{2}(d(x, z) + d(y, z)).$$

- **Menger convex** (or **M-convex**) if, for any different points $x, y \in X$, there exists a third point $z \in X$ for which $d(x, y) = d(x, z) + d(z, y)$.

- (X, d) is **metrically convex** if, for any different points $x, y \in X$ and any $\lambda \in (0, 1)$, there exists a third point $z = z(x, y, \lambda) \in X$ for which $d(x, y) = d(x, z) + d(z, y)$ and $d(x, z) = \lambda d(x, y)$.

  (X, d) is **strictly metrically convex** if the point $z(x, y, \lambda)$ is unique for all $x, y \in X$ and $\lambda \in (0, 1)$.

- (X, d) is **hyperconvex** (or **injective**) if it is metrically convex and its metric balls have the **infinite Helly property**, i.e., any family of mutually intersecting closed balls in X has non-empty intersection.

MAIN CLASSES OF METRICS

- Given a connected graph $G = (V, E)$, the **path metric** between two vertices is the number of edges of a shortest path connecting them.
- Given a finite set X and a finite set $\mathcal{O}$ of (unary) **editing operations** on X, the **editing metric** on X is the path metric of the graph with the vertex-set X and xy being an edge if y can be obtained from x by one of the operations from $\mathcal{O}$.

- On a **normed vector space** $(V, ||.||)$, the **norm metric** is $||x - y||$.
- The l_p**-metric**, $1 \leq p \leq \infty$, is $||x - y||_p$ norm metric on $\mathbb{R}^n$ (or on $\mathbb{C}^n$), where $||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ for $p < \infty$ and $||x||_\infty = \max_{1 \leq i \leq n} |x_i|$.

  The **Euclidean metric** (or **Pythagorean distance**, **as-crow-flies distance**, **beeline distance**) is l_2-metric on $\mathbb{R}^n$.
- **Banach-Mazur distance** between n-dim. **normed spaces** V and W is $\ln \inf_T \{||T|| \cdot ||T^{-1}||$, where $T : V \to W$ is an isomorphism.
- **Lipschitz distance** between metric spaces (X, d_X) and (Y, d_Y) is $\inf_f \{||f||_{Lip} \cdot ||f^{-1}||_{Lip}\}$, where infimum is over all bijective functions $f : X \to Y$ and the **Lipschitz norm** is $||f||_{Lip} = \sup\{\frac{d_Y(f(x), f(y))}{d_X(x,y)} : x, y \in X, x \neq y\}$.

- Given a **measure space** $(\Omega, \mathcal{A}, \mu)$, the **symmetric difference** (or **measure**) **semi-metric** on the set $\mathcal{A}_\mu = \{A \in \mathcal{A} : \mu(A) < \infty\}$ is $\mu(A \triangle B)$ (where $A \triangle B = (A \cup B) \backslash (A \cap B)$ is the **symmetric difference** of the sets $A, B \in \mathcal{A}_\mu$) and 0 if $\mu(A \triangle B) = 0$.

  Identifying $A, B \in \mathcal{A}_\mu$ if $\mu(A \triangle B) = 0$, gives the **measure metric**.

  If $\mu(A) = |A|$, then $|A \triangle B| = 0$ iff $A = B$ and $|A \triangle B|$ is a metric.

- Given a **measure space** $(\Omega, \mathcal{A}, \mu)$, the **Steinhaus semi-metric** on the set $\mathcal{A}_\mu = \{A \in \mathcal{A} : \mu(A) < \infty\}$ is 0 if $\mu(A \cup B) = 0$ and

$$\frac{\mu(A \triangle B)}{\mu(A \cup B)} = 1 - \frac{\mu(A \cap B)}{\mu(A \cup B)}, \text{ otherwise.}$$

  The **biotope** (or **Tanimoto**) **metric** $\frac{|A \triangle B|}{|A \cup B|}$ is the case $\mu(A) = |A|$.

The **Hamming metric** on $\mathbb{R}^n$ is $d_H = |\{i : 1 \leq i \leq n, x_i \neq y_i\}|$.

On vertices of unit cube $\{0,1\}^n$ it is l_1**-metric** and squared l_2**-metric**. Eqv., for subsets $A, B \subset X$ with $|X| = n$, it is **measure metric** $|A \triangle B|$.

The **Bray-Curtis distance** on $\mathbb{R}^n$ is $\frac{\sum |x_i - y_i|}{\sum (x_i + y_i)}$.

The **Canberra distance** on $\mathbb{R}^n$ is $\sum \frac{|x_i - y_i|}{|x_i| + |y_i|}$.

The **Mahalonobis distance** (or **statistical distance**) on $\mathbb{R}^n$ is

$$\sqrt{(detA)^{\frac{1}{n}}(x-y)A^{-1}(x-y)^T},$$

where A is a positive-definite matrix.

The **Hellinger distance** on $\mathbb{R}^n_+$ is $\sqrt{2\sum\left(\sqrt{\frac{x_i}{\overline{x}}} - \sqrt{\frac{y_i}{\overline{y}}}\right)^2}$.

Popular metrics on real plane $\mathbb{R}^2$

- The **lift** (or **raspberry picker**, **jungle river**, **barbed wire**) **metric** on $\mathbb{R}^2$ is $|x_1 - y_1|$ if $x_2 = y_2$, and $|x_1| + |x_2 - y_2| + |y_1|$ if $x_2 \neq y_2$.

- Given a **norm** $||.||$ on $\mathbb{R}^2$, the **French Metro metric** on $\mathbb{R}^2$ is $||x - y||$ if $x = cy$ for some $c \in \mathbb{R}$, and $||x|| + ||y||$, otherwise. For $||.||_2$, it is called **Paris** (or **hedgehog**, **radial**, **enhanced SNCF**) **metric**.

  In graph terms, this metric is similar to the **path metric** of the tree consisting of a point from which radiate several disjoint paths. When only one line radiates from the point, it is called **train metric**.

- For any metric space (X, d) and $f \in X$ (a **flower-shop**), the **flower-shop metric** is $d(x, f) + d(f, y)$ for $x \neq y$, and 0, otherwise. For $(X, d) = (\mathbb{R}^2, ||x - y||)$ and $f = (0, 0)$, it is called **British Rail** (or **SNCF**, **Post Office**, **caterpillar**, **shuttle**) **metric**.

- The **Moscow metric** (or **Karlsruhe metric**) between two points $x, y \in \mathbb{R}^2$ is the length of a shortest Euclidean (x, y)-path consisting only of *radial streets* (segments of straight lines passing through the origin) and *circular avenues* (segments of circles centered at the origin).

- The **Central Park metric** between two points $x, y \in \mathbb{R}^2$, is the length of a shortest **Manhattan** (x, y)**-path** at the presence of a given set of areas which are traversed by a shortest Euclidean path (for example, Central Park in Manhattan).

- Let $\mathcal{O} = \{O_1, \ldots, O_m\}$ be a collection of pairwise disjoint polygons on $\mathbb{R}^2$ representing a set of obstacles which are neither transparent, nor traversable. The **collision avoidance distance** (or **piano movers distance**) between two points $x, y \in \mathbb{R}^2 \backslash \{\mathcal{O}\}$, is the length of a shortest continuous (x, y)-path that do not intersect obstacles $O_i \backslash \partial O_i$.

Metrics on digital plane $\mathbb{Z}^2$

A **computer image** is a subset of $\mathbb{Z}^n$ (**digital nD space**). Usually, n=2. The points of $\mathbb{Z}^2$ and $\mathbb{Z}^3$ are **pixels** and **voxels**, respectively.

A **digital metric** is any integer-valued metric on a digital nD space.

Main digital metrics are: the l_1**-,** l_∞**-metrics** and (rounded to nearest, upper or lower, integer) l_2**-metric**.

A list of **neighbors** of a pixel can be seen as a list of permitted **one-step moves** on $\mathbb{Z}^2$. Associate a positive weight to each type of such move. Many digital metrics are the minimum, over all admissible paths (sequences of permitted moves) of the sum of their weights.

- The **rook metric** is a metric on $\mathbb{Z}^2$, defined as the minimum number of moves a chess rook need to travel from x to $y \in \mathbb{Z}^2$. It is $\{0, 1, 2\}$-valued and coincides with the **Hamming metric** on $\mathbb{Z}^2$.

- The **grid metric** is the l_1**-metric** on $\mathbb{Z}^n$. It is the **path metric** of an infinite graph: two points of $\mathbb{Z}^n$ are adjacent if their l_1-distance is 1. For $n = 2$, this metric is restriction on $\mathbb{Z}^2$ of **Manhattan metric** and it called 4**-metric** since each point has exactly 4 l_1-neighbors in $\mathbb{Z}^2$.

- The **lattice metric** is the l_∞**-metric** on $\mathbb{Z}^n$. It is the **path metric** of an infinite graph: two points of $\mathbb{Z}^n$ are adjacent if their l_∞-distance is 1. For $\mathbb{Z}^2$, the adjacency corresponds to the king move in chessboard terms, and this metric is called **chessboard metric** (or **king metric**, 8**-metric** since each point has exactly 8 l_∞-neighbors in $\mathbb{Z}^2$).

- The **hexagonal metric** is a metric on $\mathbb{Z}^2$ with an **unit sphere** $S^1(x)$: $S^1(x) = S^1_{l_1}(x) \cup \{(x_1 \pm 1, x_2 - 1), (x_1 \pm 1, x_2 + 1)\}$ if x_2 is odd/even.

  Since $|S^1(x)| = 6$, the hexagonal metric is called also 6**-metric**. The hexagonal metric is the **path metric** on the **hexagonal grid** of the plane. It approximates l_2-metric better than l_1- or l_∞**-metric**.

- The **knight metric** is a metric on $\mathbb{Z}^2$, defined as the minimum number of moves a chess knight would take to travel from x to $y \in \mathbb{Z}^2$.

- Let $p, q \in \mathbb{N}$ such that $p+q$ is odd, and $(p,q) = 1$.

  A (p,q)**-super-knight** (or (p,q)**-leaper**) is a (variant) chess piece a move of which consists of a leap p squares in one orthogonal direction followed by a 90 degree direction change, and q squares leap to the destination square. Chess-variant terms for an $(p,1)$-leaper with p=0, 1, 2, 3, 4: **Wazir**, **Ferz**, usual **Knight**, **Camel**, **Giraffe** and for an $(p,2)$-leaper with $p = 0, 1, 2, 3$: **Dabbaba**, **Knight**, **Alfil**, **Zebra**.

  A **super-knight metric** on $\mathbb{Z}^2$ is the minimum number of moves a (p,q)-super-knight would take to travel from x to $y \in \mathbb{Z}^2$.

  The **knight metric** is the $(1,2)$-super-knight metric.

  The l_1**-metric** is $(0,1)$-super-knight metric, i.e., the **Wazir metric**.

- Given $\alpha, \beta \geq 0$ with $\alpha \leq \beta < 2\alpha$, consider (α, β)**-weighted l_∞-grid**, i.e., pixel graph $(V = \mathbb{Z}^2, E)$ with $(xy) \in E$ if $|x - y|_\infty = 1$, and horizontal/vertical and diagonal edges having **weights** α and β, resp. Borgefors (α, β)**-chamfer metric** is the **weighted path metric** of this graph. The main cases are (α, β)=$(1, 0)$ (l_1**-metric**), $(3, 4)$, $(1, 1)$ (l_∞**-metric**), $(1, \sqrt{2})$ (**Montanari metric**), $(5, 7)$ (**Verwer metric**), $(2, 3)$ (**Hilditch-Rutovitz metric**).

- An (α, β, γ)**-chamfer metric** is the weighted path metric of voxel graph $(V = \mathbb{Z}^3, E)$ with $(xy) \in E$ if $|x - y|_\infty = 1$, and moves to 6 face, 12 edge, 8 corner neighbors having **weights** α, β, γ, respectively.

  The cases (α, β, γ)=$(1, 1, 1)$ (l_∞**-metric**), $(3, 4, 5)$, $(1, 2, 3)$ are the most used ones for digital $3D$ images.

AUDIO DISTANCES

Sound: vibration of air particles causing pressure variations in eardrums.

Audio (speech, music, etc.) **Signal Processing** is the processing of analog (continuous) or, mainly, digital representation of the air pressure waveform of the sound. A **sound spectrogram** (or **sonogram**) is a visual 3D representation of an acoustic signal. It is obtained either by series of bandpass filters (an analog processing), or by application of the **short-time Fourier transform** to the electronic analog of an acoustic wave.

Three axes represent time, frequency and **intensity**. Often this 3D curve is reduced to 2D by indicating the intensity with, say, more thick lines.

Sound is called **tone** if it is periodic (the lowest **fundamental** frequency plus its multiples, **harmonics**) and **noise**, otherwise. The frequency is measured in **cps** (the number of complete cycles per second) or Hz (Herzs). The range of audible sound frequencies to humans is 20Hz–18kHz.

Decibel dB is the unit used to express relative strength of two signals. Audio signal's amplitude in dB is $20 \log_{10} \frac{A(f)}{A(f')} = 10 \log_{10} \frac{P(f)}{P(f')}$, where f' is a reference signal selected to correspond 0 dB (human hearing threshold). The threshold of pain is about $120 - 140$ dB.

Pitch and **loudness** are psycho-acoustic (auditory subjective) terms for frequency and amplitude.

Mel scale correspond to the auditory sensation of tone height and based on **mel**, a unit of pitch (perceived frequency). It is connecccted to the acoustic frequency f Hz scale by $Mel(f) = 1127 \ln(1 + \frac{f}{700})$.

Psycho-acoustic **Bark scale** of loudness range from 1 to 24 corresponding to the first 24 critical bands of hearing $(0, 100, \ldots, 12000, 15500$ Hz).

Those bands correspond to spatial regions of the basilar membrane of the inner ear, where oscillations produced by the sound activate the hair cells and neurons. $Bark(f) = 13 \arctan(0.76f) + 3.5 \arctan(\frac{f}{0.75})^2$ in f kHz scale.

Human **phonation** (speech, song, laughter) is controlled usually by **vocal tract** (the throat and mouth) shape. This shape, i.e., the cross-sectional profile of the tube from the closure in the **glottis** (the space between the vocal cords) to the opening (lips), is represented by the cross-sectional area function $Area(x)$, where x is the distance to glottis.

The vocal tract acts as a resonator during vowel phonation, because it is kept relatively open. Those resonances reinforce the source sound (ongoing flow of lung air) at particular **resonant frequencies** (or **formants**) of the vocal tract, producing peaks in the **spectrum** of the sound.

Each **vowel** has two characteristic formants, depending of the vertical and horizontal position of the tongue in the mouth.

The frequency of speech signal ($3 - 8$ Hz) resonates with the theta rhythm of neocortex. Speakers produce $3 - 8$ syllabes per second.

The **spectrum** of a sound is the distribution of magnitude (dB) of the components of the wave. The **spectral envelope** is a smooth contour connecting spectral peaks. Estimation of the spectral envelopes is based on either LPC (linear predictive coding), or FTT (fast Fourier transform).

FTT maps time-domain functions into frequency-domain. The **cepstrum** of the signal $f(t)$ is $FT(\ln(FT(f(t) + 2\pi mi)))$, where m is the integer needed to unwrap the angle or imaginary part of the complex log function.

(The complex and real cepstrum use, respectively, complex and real log function. The real cepstrum uses only the magnitude of the original signal $f(t)$, while the complex cepstrum uses also phase of $f(t)$.)

FFT performs Fourier transform on the signal and sample the discrete transform output at the desired frequencies in mel scale.

Power spectral density $PSD(f)$ of a wave is the power per Hz. It is the Fouriet transform of the autocorrelation sequence. So, the power in the band $(-W, W)$ is $\int_{-W}^{W} PSD(f)df$. A **power-law noise** has $PSD(f) \sim f^{\alpha}$.

The noise is called **violet**, **blue**, **white**, **pink** (or $\frac{1}{f}$ **noise**), **red** (or **brown(ian)**), **black** (or **silent**) if $\alpha = 2, 1, 0, -1, -2, < -2$, respectively.

PSD changes by 3α dB per **octave** (distance between a frequency and its double); it decreases for $\alpha < 0$.

Pink noise occurs in many physical, biological and economic systems. It has equal power in proportionally wide frequency ranges.

Humans also process frequencies in a such logarithmic space (approximated by the **Bark scale**). Every octave contains the same amount of energy, and pink noise is used as a reference signal in audio engineering. Steady pink noise (incl. light music) reduces brain wave complexity and improve sleep.

Parameter-based distances used in recognition and processing of speech data are usually derived by LPC, modeling speech spectrum as a linear combination of the previous samples (as in autoregressive process).

Majority of **distortion measures between sonograms** are variations of **squared Euclidean distance** (including **Mahalanobis distance**) and probabilistic distances (f**-divergence of Csizar**, **Chernoff distance**, generalized **total variation metric**).

The distances for sound processing below are between vectors x and y representing two signals to compare.

For **recognition**, they are a template reference and input signal, while for **noise reduction**, they are original (reference) and distorted signal.

Often distances are calculated for small segments, between vectors representing short-time spectra, and then averaged.

- Given a sound, let P and A_s denote its average power and RMS (root mean square) amplitude. The **signal-to-noise ratio in decibels** is

$$10\log_{10}(\frac{P_{signal}}{P_{noise}}) = P_{signal,dB} - P_{noise,dB} = 10\log_{10}(\frac{A_{signal}}{A_{noise}})^2.$$

  The **dynamical range** is such ratio between the strongest undistorted and minimum discernable signals. The **Shannon-Hartley theorem** express the **capacity** (maximal possible information rate) of a channel with additive **colored** (frequency-dependent) Gaussian noise, on the bandwidth B in Hz as $\int_0^B \log_2(1 + \frac{P_{signal}(f)}{P_{noise}(f)})df$.

  The **SNR distance** between signals $x = (x_i)$ and $y = (y_i)$ with n frames is $10\log_{10}\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - y_i)^2}$.

- The **segmented SNR** is $\frac{10}{m}\sum_{m=0}^{M-1}\left(\log_{10}\sum_{i=nm+1}^{nm+n}\frac{x_i^2}{(x_i-y_i)^2}\right)$.

- The **Czekanovski-Dice distance** is $\frac{1}{n}\sum_{i=1}^n\left(1 - \frac{2\min\{x_i,y_i\}}{x_i+y_i}\right)$.

- **Spectral distances**

  Given two discrete spectra $x = (x_i)$ and $y = (y_i)$ with n channel filters, their **Euclidean metric** EM, **slope metric** SM (Klatt, 1982) and **2nd differential metric** $2DM$ (Assmann and Summerfield, 1989) are defined, respectively, by

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2},\ \sqrt{\sum_{i=1}^{n}(x'_i - y'_i)^2} \text{ and } \sqrt{\sum_{i=1}^{n}(x''_i - y''_i)^2},$$

  where $z'_i = z_{i+1} - z_i$ and $z''_i = \max(2z_i - z_{i+1} - z_{i-1}, 0)$.

  Comparing, say, the auditory excitation patterns of vowels, EM gives equal weight to peaks and troughs although spectral peaks have more perceptual weight. SM emphasizes the formant frequencies, while $2DM$ sets to zero the spectral properties other than the formants.

The **RMS log spectral distance** (or **root-mean-square distance**, **mean quadratic distance**) $LSD(x, y)$ is defined by $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \ln y_i)^2}$. The corresponding l_1- and l_∞-distances are called **mean absolute distance** and **maximum deviation**. These three distances are related to decibel variations in the log spectral domain by the multiple $\frac{10}{\log 10}$.

$LSD^2(x, y)$, via the cepstrum representation $\ln x(\omega) = \sum_{j=-\infty}^{\infty} c_j e^{-j\omega i}$ (where $x(\omega)$ is the **power cepstrum** $|FT(\ln(|FT(f(t))|^2)|^2)$ becomes, in the complex cepstral space, the **cepstral distance**.

The **log area ratio distance** $LAR(x, y)$ between x and y is defined by $\sqrt{\frac{1}{n}\sum_{i=1}^{n} 10(\log_{10} Area(x_i) - \log_{10} Area(y_i))^2}$, where $Area(z_i)$ is the cross-sectional area of the i-th segment of the vocal tract.

- The **spectral magnitude-phase distortion** between signals $x = x(\omega)$ and $y = y(\omega)$ is $\frac{1}{n}\left(\lambda \sum_{i=1}^{n}(|x(w)| - |y(w)|)^2 + (1-\lambda)\sum_{i=1}^{n}(\angle x(w) - \angle y(w))^2\right)$, where $|x(w)|$, $|y(w)|$ are magnitude spectra, and $\angle x(w)$, $\angle y(w)$ are phase spectra of x and y, resp, while parameter $\lambda, 0 \leq \lambda \leq 1$, is scaling factor to attach commensurate weights to the magnitude and phase terms.

  The case $\lambda = 0$ corresponds to the **spectral phase distance**.

  Given a signal $f(t) = ae^{-bt}u(t)$, $a, b > 0$, which has Fourier transform $x(w) = \frac{a}{b+iw}$, its **magnitude** (or **amplitude**) spectrum is $|x| = \frac{a}{\sqrt{b^2+w^2}}$, and its **phase** spectrum (in radians) is $\alpha(x) = \tan^{-1}\frac{w}{b}$, i.e., $x(w) = |x|e^{i\alpha} = |x|(\cos\alpha + i\sin\alpha)$.

- The **Bark spectral distance** is a perceptual distance $BSD(x,y) = \sum_{i=1}^{n}(x_i - y_i)^2$, i.e., is the **squared Euclidean distance** between **Bark spectra** (x_i) and (y_i) of x and y, where i-th component corresponds to i-th auditory critical band in Bark scale.

- The **Itakura-Saito quasi-distance** (or **maximum likelihood distance**) $IS(x,y)$ between LPC-derived spectral envelopes $x = x(\omega)$ and $y = y(\omega)$ is $\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\ln\frac{x(w)}{y(w)} + \frac{y(w)}{x(w)} - 1\right)dw$.

  The **cosh distance** is defined by $IS(x,y) + IS(y,x)$.

- The **log likelihood ratio quasi-distance** (or **Kullback-Leibler distance**) $KL(x,y)$ between LPC-derived spectral envelopes $x = x(\omega)$ and $y = y(\omega)$ is defined by $\frac{1}{2\pi}\int_{-\pi}^{\pi} x(w)\ln\frac{x(w)}{y(w)}dw$. The **Jeffrey divergence** $KL(x,y) + KL(y,x)$ is also used.

"Quefrency", "cepstrum": anagrams of "frequency", "spectrum", resp.

- The **RMS log spectral distance** (or **root-mean-square distance**) $LSD(x, y)$ between discrete spectra $x = (x_i)$ and $y = (y_i)$ is Euclidean distance $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \ln y_i)^2}$. The square of it, via cepstrum representation $\ln x(\omega) = \sum_{j=-\infty}^{\infty} c_j e^{-ij\omega}$ is the **cepstral distance**.

- The **cepstral distance** (or **squared Euclidean cepstrum metric**) $CEP(x, y)$ between LPC-derived spectral envelopes $x = x(\omega)$ and $y = y(\omega)$ is $\frac{1}{2\pi}\int_{-\pi}^{\pi}(\ln x(w) - \ln y(w))^2\,dw = \sum_{j=-\infty}^{\infty}(c_j(x) - c_j(y))^2$, where $c_j(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iwj}\ln|z(w)|dw$ is j-th cepstral (real) coefficient of z derived by Fourier transform or LPC.

  The **quefrency-weighted cepstral distance** (or **weighted slope distance**) between x and y is $\sum_{i=-\infty}^{\infty} i^2(c_i(x) - c_i(y))^2$.

  The **Martin cepstrum distance** between two AR (autoregressive) models is, in terms of their cepstrums, $(\sum_{i=0}^{\infty} i(c_i(x) - c_i(y))^2)^{\frac{1}{2}}$.

- In **Poetry, meter** (or *cadence*) is a measure of rhythmic quality, the regular linguistic sound patterns of verse. *Qualitative meter* indicate syllables coming at regular intervals. *Mono*-, *di*-, *trimeter*, etc. indicate the number of *feet* (specific sequences of syllable types).

- In **Music, meter** (or *metre*) is the regular rhythmic patterns of musical line, the division of a composition into parts of equal time, and the subdivision of them. It is derived from the poetic meter of song. Different tonal preferences in music are closely related to the differences in the tonal characteristics of voiced speech.

  Isometre is the use of *pulse* (unbroken series of periodically occurring short stimuli) without regular meter, and *polymetre* is the use of two or more different meters simultaneously whereas *multimetre* is using them in succession. A rhythmic pattern or unit is either *intrametric* (confirming the pulses on the metric level), or *contrametric* (syncopated, not following the meter), or *extrametric* (irregular).

- A rhythm timeline (music pattern) is represented, besides standard music notation, as binary vector, pitch vector, pitch difference vector, chronotonic histogram or, for example as:

  1. a **inter-onset interval vector** $t = (t_1, \ldots, t_n)$ of n time intervals between consecutive onsets.

  2. a **rhythm difference vector** $r = (r_1, \ldots, r_{n-1})$, where $r_i = \frac{t_{i+1}}{t_i}$.

  Examples of general **distances between rhythms** are Hamming distance, **swap metric**, **Earth Mover distance** between their given vector representations. The **Euclidean interval vector distance** is the Euclidean distance between two inter-onset interval vectors.

  Coyle-Shmulevich **interval-ratio distance** is $1 - n + \sum_{i=1}^{n-1} \frac{\max\{r_i, r'_i\}}{\min\{r_i, r'_i\}}$, where r and r' are rhythm difference vectors of two rhythms.

- **Pitch** is a subjective correlate of the fundamental frequency A **musical scale** is a linearly ordered collection of pitches (notes).

  A **pitch distance** (or **interval**, **musical distance**) is the size of the section of the linearly-perceived pitch-continuum bounded by those two pitches, as modeled in a given scale. The pitch distance between two successive notes in a scale is called a **scale step**.

  In Western music now, the most used one is the **chromatic scale** (octave of 12 notes) of **equal temperament**, i.e., divided into 12 equal steps with the ratio $\sqrt[12]{2}$ between any two adjacent frequencies. The scale step here is a **semitone**, i.e., the distance between two adjacent keys (black and white) on a piano. The **distance between notes** whose frequencies are f_1 and f_2 is $12\log_2(\frac{f_1}{f_2})$ semitones.

  A MIDI (Musical Instrument Digital Interface) number of fundamental frequency f is $p(f) = 69 + 12\log_2\frac{f}{440}$. In terms of MIDI numbers, the distance between notes is the **natural metric** $|m(f_1) - m(f_2)|$.

- **Spatial music** is electroacoustic music and sound art in which the location and movement of sound sources, in physical or virtual space, is a primary compositional parameter and a central feature.

  Space music is gentle, harmonious sound that facilitates the experience of contemplative spaciousness. Generating serenity and imagination, it is associated with ambient, New Age, and electronic music.

- **Long-distance drumming** (or *drum telegraphy*) is an early form of long-distance communication which was used by cultures in Africa, New Guinea and the tropical America living in deforested areas. A rhythm could represent an signal or simply be subject to musical laws.

  The *message drums* (or *slit gongs*) were developed from hollow tree trunks. The sound could be understood at ≤ 8 km but usually it was relayed to a next village.

  Another oldest tools of audio telecommunication were *horns*.

- **Sonority distance effect**

  People in warm-climate cultures spend more time outdoors and engage, on average, in more distal oral communication. Such populations have greater sonority (audibility) of their phoneme inventory and speakers use more simple consonant-vowel syllables, vowels and *sonorant* (say, nasal "n", "m" rather than obstruents as "t", "g") consonants.

  Ember, 2007: more cold months and sparse vegetation predicts less sonority. Larger mean distance of the baby from its caregivers and higher frequency of premarital/extramarital sex predicts more sonority.

  Lomax, 1968: sexual inhibition discourages speaking with a wide open mouth and predicts, in folk song style: less vocal width (ranging from singing with a very pinched, narrow, squeezed voice to the very wide and open-throated singing tone of Swiss yodelers) and greater nasality (when the sound is forced through the nose).

- **Acoustics distances**

  The *wavelength* is the distance the sound wave travels to complete one cycle. This distance is measured perpendicular to the wavefront in the direction of propagation between one peak of a *sine wave* (sinusoid) and the next corresponding peak. The wavelength of any frequency may be found by dividing the speed of sound (331.4 m/s at sea level) in the medium by the fundamental frequency.

  The *near field* is the part of a sound field (usually within about two wavelengths from the source) where there is no simple relationship between sound level and distance. The **far field** is the area beyond the near field boundary. It is comprised of the *reverberant field* and *free field*, where sound intensity decreases as $\frac{1}{d^2}$ with the distance d from the source. This law corresponds to a reduction of ≈ 6 dB in the sound level for each doubling of distance and to halving of loudness (subjective response) for each reduction of ≈ 10 dB.

- The **critical distance** (or *room radius*) is the distance from the source at which the direct sound and reverberant sound (reflected echo produced by the direct sound bouncing off, say, walls, floor, etc.) are equal in amplitude.

  The **proximity effect (audio)** is the anomaly of low frequencies being enhanced when a directional microphone is very close to the source.

  Auditory **distance cues** are based on differences in loodness, spectrum, direct-to-reverb ratio and binaural ones.

  The closer sound object is looder, has more bass, high-frequencies, transient detail, dynamic contrast. Also, it appear wider, has more direct sound level over its reflected sound and has greater time delay between the direct sound and its reflections.

- The *acoustic metric* is the term used occasionally for some distances between vowels; for example, the Euclidean distance between vectors of formant frequencies of pronounced and intended vowel.

  In Acoustics and in Fluid Dynamics, the **acoustic metric** (or **sonic metric**) is a characteristic of sound-carrying properties of a given medium: air, water, etc.

  In General Relativity and Quantum Gravity, it is a characteristic of signal-carrying properties in a given *analog model* (with respect to Condensed Matter Physics).

  When the fluid's speed becomes supersonic, the sound waves cannot come back, i.e., there is a *mute hole*, the acoustic analog of a *black hole*.

- The **cosmic light horizon** (or **Hubble radius**) is an increasing distance $D_H = ct_H$ that a light signal could have traveled since the Big Bang. Now $t_H \approx 13.7$ billion years, and $D_H \approx 13.7$ billion light-years.

  Baryon acoustic oscillations started at t=0 (post-inflation) and stopped at t=t_r (recombination). The **cosmic sound horizon** is the distance sound waves have traveled. At r, it was $\approx c_s t_r \sim 100$ kpc, now: $120 - 150$ Mpc.

  Cosmic background radiation (CMB) is thermal radiation filling the Universe almost uniformly. It originated $t_r \approx 379,000$ years after the Big Bang at **recombination**, when the Universe (ionized plasma of electrons and *baryons*, i.e., protons+neutrons) cooled to < 3000 K.

  The electrons and protons start to form neutral hydrogen atoms, allowing photons to travel freely. During next $\approx 100,000$ years radiation decoupled from matter and the Universe became transparent.

The plasma of photons and baryons can be seen as a single fluid. The gravitational collapse around "seeds" (point overdensities produced during inflation) into dark matter hierarchical halos was opposed by outward heat radiation pressure from the photon-matter interactions. This competition created longitudinal (acoustic) oscillations in the photon-baryon fluid, analogic to sound waves, created in air by pressure differences.

At recombination, the only remaining force on baryons is gravitation, and the pattern of oscillations (configuration of baryons and, at the centers of perturbations, dark matter) became frozen into the CMB. Baryon cooling into gas and stars let this pattern to grow into structure of the Universe.

More matter existed at the centers and edges of these waves, leading to to more galaxies there eventually. We detect the sound waves (periodic fluctuations in the density of the visible matter) via CMB anisotropies.

- **Sound attenuation with distance**

  Vibrations propagate through elastic solids and liquids, including the Earth, and consist of *elastic* (or *seismic*, *body*) waves and surface waves. Elastic waves are: primary (P) wave moving in the propagation direction of the wave and *secondary* (S) wave moving in this direction and perpendicular to it. Also, because the surface acts as an interface between solid and gas, surface waves occur:

  the *Love* wave moving perpendicular to the direction of the wave and the *Rayleigh* (R) wave moving in the direction of the wave and circularly within the vertical surface perpendicular to it.

  The geometric attenuation of P- and S-waves is proportional to $\frac{1}{d^2}$, when propagated by the surface of an infinite elastic body, and it is proportional to $\frac{1}{d}$, when propagated inside it.

  For the R-wave, it is proportional to $\frac{1}{\sqrt{d}}$.

Sound propagates through gas (say, air) as a P-wave. It attenuates geometrically over a distance, normally at a rate of $\frac{1}{d^2}$: the inverse-square distance law relating the growing radius d of a wave to its decreasing intensity. The **far field** is the part of a sound field in which sound pressure decreases as $\frac{1}{d}$ but its intensity decreases as $\frac{1}{d^2}$.

In natural media, further weakening occurs from *attenuation*, i.e., *scattering* (reflection of the sound in other directions) and *absorption* (conversion of the sound energy to heat).

The **sound extinction distance** is the distance over which its intensity falls to $\frac{1}{e}$ of its original value. For sonic boom intensities (say, supersonic flights), the lateral extinction distance is the distance where in 99% of cases the sound intensity is lower than $0.1 - 0.2$ mbar ($10 - 20$ pascals) of atmospheric pressure.

Water is transparent to sound. Sound energy is absorbed (due to viscosity) and $\approx 6\%$ of it scattered (due to water inhomogeneities). Sound attenuation by zoo-plankton is used in hydroacoustic measurement of fish and zoo-plankton abundance.

Absorbed less in liquids and solids, low frequency sounds can propagate in these media over much greater distances along lines of minimum sound speed (**SOFAR channel**).

On the other hand, high frequency waves attenuate more rapidly. So, low frequency waves are dominant further from the source (say, a musical band or earthquake).

Attenuation of ultrasound waves with frequency f MHz at a given distance r cm is $\alpha f r$ decibels, where α is the *attenuation coefficient* of the medium. It is used in Ultrasound Biomicroscopy; in a homogeneous medium (so, without scattering) α is 0.0022, 0.18, 0.85, 20, 41 for water, blood, brain, bone, lung, respectively.

- **Animal long-distance communication**

  The main modes of animal communication are infrasound (< 20 Hz), sound, ultrasound (> 20 kHz), vision (light), chemical (odor), tactile and electrical. Infrasound, low-pitched sound (as territorial calls) and light in air can be long-distance.

  A blue whale infrasound can travel thousands of kilometers through the ocean water using SOFAR channel (a layer where the speed of sound is at a minimum, because water pressure, temperature, and salinity cause a minimum of water density.

  Janik, 2000, estimated that unmodulated dolphin whistles at 12 kHz in a habitat having a uniform water depth of 10 m would be detectable by conspecifics at distances of $1.5 - 4$ km.

  Many animals hear infrasound of earthquakes, tsunami and hurricanes before they strike. Elephants can hear storms $160 - 240$ km away.

Most elephant communication is in the form of infrasonic rumbles which may be heard by other elephants at 5 km away and, in optimum atmospheric conditions, at 10 km. The resulting seismic waves can travel $16 - 32$ km through the ground.

But non-fundamental harmonics of elephant calls are sonic. McComb-Reby-Baker-Moss-Sayialel, 2003, found that, for female African elephants, the peak of social call frequency is ≈ 115 Hz and the *social recognition distance* (over which a contact call can be identified as belonging to a family) is usually 1.5 km and at most 2.5 km.

High-frequency sounds attenuate more rapidly with distance; they are more directional and vulnerable to scattering. But ultrasounds are used by bats (echolocation) and antropods. Rodents use them to communicate selectively to nearby receivers without alerting predators and competitors. Some anurans shift to ultrasound signals in the presence of continuous background noise (as waterfall, human traffic).

- A **phone** is a sound segment that possess distinct acoustic properties, the basis sound unit. A **phoneme** is a minimal distinctive feature/unit (a set of phones perceived as equivalent in a given language).

  The number of phonemes range, among about 6000 spoken now languages, from 11 in Rotokas to 112 in Taa (languages spoken by ≈ 4000 people in Papua New Guinea and Botswana, respectively.) Pirahã (Amazon's tribe) require gender difference in pronunciation: men use larger articulatory space and only men use the phoneme "s".

  Two main classes of **phone distance** between phones x and y are: **Spectrogram-based distances**: physical-acoustic distortion measures between the sound spectrograms of x and y;

  Feature-based phone distances: usually **Manhattan distance** $\sum_i |x_i - y_i|$ between vectors (x_i) and (y_i) representing phones x and y with respect to given inventory of phonetic features (for example, nasality, stricture, palatalization, rounding, sillability).

- The **Laver consonant distance** refers, for 22 consonantal phonemes of English, the improbability of confusing them, developed by Laver, 1994, from subjective auditory impressions.

  The smallest distance, 15%, is between $[p]$ and $[k]$, the largest one, 95%, is, for example, between $[p]$ and $[z]$. Laver also proposed a quasi-distance based on the likehood that one consonant will be misheard as another by an automatic speech-recognition system.

- Liljencrans and Lindlom, 1972, developed a **vowel space** of 14 vowels. Each vowel, after a procedure maximizing contrast among them, is represented by a pair (x, y) of resonant frequencies of the vocal tract (1st and 2nd formants) in linear mel units with $350 \leq x \leq 850$ and $800 \leq y \leq 1700$). Higher x values correspond to lower vowels and higher y values to less rounded or farther front vowels. For example, $[u]$, $[a]$, $[i]$ are represented by $(350, 800)$, $(850, 1150)$, $(350, 1700)$, resp.

- The **phonetic word distance** between two words x and y is the cost-based **editing metric** (for phone sustitutions and indels).

  A word is seen as a string of phones. Given a **phone distance** $r(u, v)$ on the International Phonetic Alphabet with additional phone 0 (the silence), the cost of substitution of phone u by v is $r(u, v)$, while $r(u, 0)$ is the cost of insertion or deletion of u.

- The **linguistic distance** (or **dialectal distance**) between language varieties X and Y is the mean, for fixed sample S of notions, **phonetic word distance** between **cognate** (i.e., having the same meaning) words s_X and s_Y, representing the same notion $s \in S$ in X and Y, respectively. Cf. Dutch-German *dialect continuum*.

- **Stover's distance** between phrases with the same key word is the sum $\sum_{-n \leq i \leq +n} a_i x_i$, where $0 < a_i < 1$ and x_i is the proportion of non-mathched words between the phrases within a moving window.

- The main phonetic encoding algorithms are (based on English language pronunciation) *Soundex*, *Phonix* and *Phonex*, converting words into one-letter three-digits codes. The letter is the first one in the word and the three digits are derived using an assignment of numbers to other word letters. Soundex and Phonex assign:

  0 to a, e, h, i, o, u, w, y; 1 to b, p, f, v; 2 to c, g, j, k, q, s, x, z; 3 to d, t; 4 to l; 5 to m, n; 6 to r.

  The **Editex distance** (Zobel-Dart, 1996) between two words x and y is a cost-based **editing metric** (i.e., the minimal cost of transforming x into y by substitution, deletion and insertion of letters). The costs for substitutions, are 0 if two letters are the same, 1 if they are in the same letter group, and 2, otherwise.

  The *syllabic alignment distance* (Gong-Chan, 2006) between two words x and y is another cost-based **editing metric**. It is based on Phonix.

IMAGE DISTANCES

Image Processing treat signals such as photographs, video, or tomographic output. In particular, **Computer Graphics** consists of image synthesis from some abstract models, while **Computer Vision** extracts some abstract information. From ≈ 2000: mainly digitally.

Computer graphics (and our brains) deals with **vector graphics** images, i.e., those represented geometrically by curves, polygons, etc. A **raster graphics image** (or **digital** image) is a representation of $2D$ image as a finite set of digital values, **pixels**, on square ($\mathbb{Z}^2$) grid.

Video and tomographic (MRI) images are $3D$ ($2D$ plus time).

A **digital binary** image corresponds to only two values 0,1 with 1 being interpreted as logical "true" and displayed as black. A **binary continuous** image is a compact subset of Euclidean space $\mathbb{E}^n$, n=2, 3

The **gray-scale images** can be seen as point-weighted binary images. In general, a **fuzzy set** is a point-weighted set with weights (**degrees of membership**. **Histogram** of a a gray-scale image gives the frequency of brightness values in it.

Humans can differ between ≈ 350000 colors but only 30 gray-levels.

For **color images**, (RGB)-representation is most known, where space coordinates R, G, B indicate red, green and blue level.

Among other color models (spaces) are: (CMY) cube (Cyan, Magenta, Yellow colors), (HSL) cone (Hue-color type given as angle, Saturation in %, Luminosity in %), and (YUV), (YIQ) used in PAL, NTSC TV.

(RGB) converts into gray-level luminance by $0.299R + 0.587G + 0.114B$

A **color space** is a 3-parameter description of colors. Exactly 3 are needed because 3 kinds of receptors (cells on the retina) exist in the human eye: for short, middle, long wavelengths, i.e., blue, green, red.

The basic assumption of Colorimetry is that the perceptual color space admits a metric, the true **color distance**. It is expected to be locally Euclidean, i.e., a **Riemannian metric**. Another assumption: there is a continuous mapping from this metric to the one of light stimuli.

Probability-distance hypothesis: the probability with which one stimulus is discriminated from another is a (continuously increasing) function of some subjective quasi-metric between these stimuli.

Such **uniform color scale**, where equal distances in the color space correspond to equal differences in color, is not obtained yet and existing **color distances** are various approximations of it.

Images are often represented by **feature vectors**, including color histograms, color moments, textures, shape descriptors, etc.

Examples of feature (parameter) spaces are:

raw intensity (pixel values), **edges** (contours, boundaries, surfaces), **salient features** (corners, line intersections, points of high curvature), and **statistical features** (moment invariants, centroids). Typical video features are in terms of overlapping frames and motions.

Image Retrieval (similarity search) consists of (as for **pattern recognition** with other data: audio, DNA/protein sequences, text documents, time series etc.) finding images whose features values are similar either between them, or to given query or in given range.

Distances are between, for Image Retrieval, feature vectors of a query and reference, and, for Image Processing (as Audio Noise Reduction), approximated and "true" digital images (to evaluate algorithms).

There are two methods to compare images directly (without features): intensity-based (color and texture histograms) and geometry-based (shape representations as **medial axis**, **skeletons**).

Unprecise term **shape** is used for the extent (silhouette) of the object, for its local geometry or geometrical pattern (conspicuous geometric details, points, curves, etc.), or for that pattern modulo a similarity transformation group (translations, rotations, and scalings).

Unprecise term **texture** means all what is left after color and shape have been considered, or it is defined via structure and randomness.

The similarity between vector representations of images is measured usually by l_p**-**, **weighted editing**, **probabilistic** distances, etc.

The main distances used for compact subsets X and Y of $\mathbb{R}^n$ (usually, $n = 2, 3$) or their digital versions are: **Asplund**, **Shephard metrics**, $vol(X \Delta Y)$ and variations of the **Hausdorff distance**.

Riemannian color space: the proposal to measure perceptual dissimilarity of colors by a **Riemannian metric** on a convex cone $C \subset \mathbb{R}^3$ comes from von Helmholtz, 1892, and Luneburg, 1947.

The only such *GL-homogeneous* cones C (i.e., the group of orientation-preserving linear self-maps of $\mathbb{R}^3$ acts transitively on C) are either $C_1 = \mathbb{R}_{>0} \times (\mathbb{R}_{>0} \times \mathbb{R}_{>0})$, or $C_2 = \mathbb{R}_{>0} \times C'$, where C' is the set of 2×2 real symmetric matrices with determinant 1. The first factor $\mathbb{R}_{>0}$ can be identified with variation of brightness and the other with the set of lights of a fixed brightness. Let $\alpha_i > 0$ be some constants.

The **Stiles color metric** (Stiles, 1946) is the GL-invariant Riemannian metric on $C_1 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_i > 0\}$ given by the **line element** $ds^2 = \sum_{i=1}^3 \alpha_i (\frac{dx_i}{x_i})^2$.

The **Resnikoff color metric** (Resnikoff, 1974) is the GL-invariant Riemannian metric on $C_2 = \{(x, u) : x \in \mathbb{R}_{>0}, u \in C'\}$ given by the **line element** $ds^2 = \alpha_1 (\frac{dx}{x})^2 + \alpha_2 ds^2_{C'}$, where $ds^2_{C'}$, is the **Poincaré metric**

- For a given $3D$ color space and a list of n colors, let (c_{i1}, c_{i2}, c_{i3}) be the representation of the i-th color of the list in this space.

  For a color histogram $x = (x_1, \ldots, x_n)$, its **average color** is the vector $(x_{(1)}, x_{(2)}, x_{(3)})$, where $x_{(j)} = \sum_{i=1}^{n} x_i c_{ij}$ (for example, the average red, blue and green values in (RGB)).

  The **average color distance** between two color histograms is the Euclidean distance of their average colors.

- Given an image (as a subset of $\mathbb{R}^2$), let p_i be the area percentage of it occupied by the color c_i. A **color component** of the image is (c_i, p_i).

  The **Ma-Deng-Manjunath distance** between color components (c_i, p_i) and (c_j, p_j) is $|p_i - p_j| \cdot d(c_i, c_j)$, where $d(c_i, c_j)$ is the distance between colors c_i and c_j in a given color space.

- Given two color histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ (with x_i, y_i representing number of pixels in the bin i), their Swain-Ballard's **histogram intersection quasi-distance** is $1 - \frac{\sum_{i=1}^{n} \min\{x_i, y_i\}}{\sum_{i=1}^{n} x_i}$.

  For normalized histograms (total sum is 1) above quasi-distance is the usual l_1**-metric** $\sum_{i=1}^{n} |x_i - y_i|$. Their Rosenfeld-Kak's **normalized cross correlation** is a similarity $\frac{\sum_{i=1}^{n} x_i, y_i}{\sum_{i=1}^{n} x_i^2}$.

- Given two color histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ (usually, $n = 256$ or $n = 64$) representing the color percentages of two images, their **histogram quadratic distance** is **Mahalonobis distance**, defined by $\sqrt{(x-y)^T A (x-y)}$, where $A = ((a_{ij}))$ is a symmetric positive-definite matrix, and weight a_{ij} is some, perceptually justified, similarity between colors i and j.

  For example, $a_{ij} = 1 - \frac{d_{ij}}{\max_{1 \le p,q \le n} d_{pq}}$, where d_{ij} is the Euclidean distance between 3-vectors representing i and j in some color space.

- Given two histogram-based descriptors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, their **histogram diffusion distance** (Ling-Okada, 2006) is defined by

$$\int_0^T ||u(t)||_1 dt,$$

where T is a constant, and $u(t)$ is a heat diffusion process with initial condition $u(0) = x - y$. In order to approximate the diffusion, the initial condition is convoluted with a Gaussian window; then the sums of l_1-norms after each convolution approximate the integral.

- Let $f(x)$ and $g(x)$ denote brightness values of two digital gray-scale images f and g at the pixel $x \in X$, where X is a raster of pixels. Any distance between point-weighted sets (X, f) and (X, g) can be applied as **gray-scale image distance** between f and g. The main used ones:

  RMS (**root mean-square error**) $\left(\frac{1}{|X|}\sum_{x\in X}(f(x)-g(x))^2\right)^{\frac{1}{2}}$;

  Signal-to-noise ratio $SNR(f,g) = \left(\frac{\sum_{x\in X} g(x)^2}{\sum_{x\in X}(f(x)-g(x))^2}\right)^{\frac{1}{2}}$;

  Pixel misclassification error rate $\frac{1}{|X|}|\{x \in X : f(x) \neq g(x)\}|$;

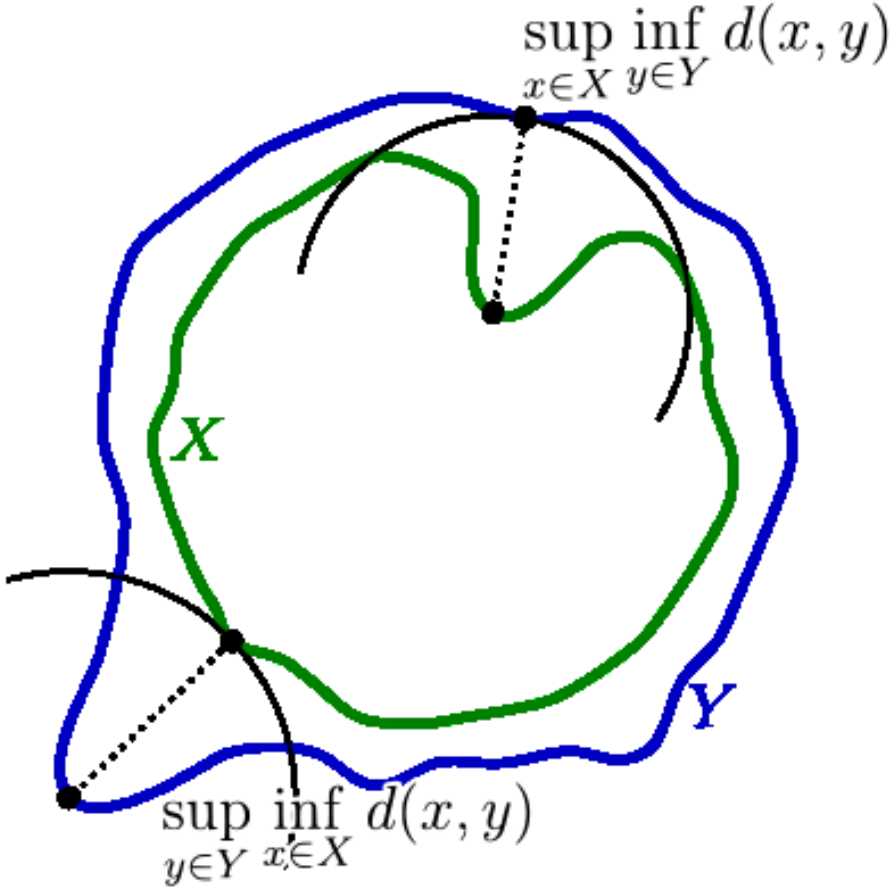
  Frequency RMS $\left(\frac{1}{|U|^2}\sum_{u\in U}(F(u)-G(u))^2\right)^{\frac{1}{2}}$, where F, G are the discrete Fourier transforms of f, g, and U is the frequency domain;

  Sobolev norm of order δ error $\left(\frac{1}{|U|^2}\sum_{u\in U}(1+|\eta_u|^2)^{\delta}(F(u)-G(u))^2\right)^{\frac{1}{2}}$, where $0 < \delta < 1$ is usually $\frac{1}{2}$), and η_u is the $2D$ frequency vector associated in U with position u.

- Given a number r, $0 \leq r < 1$, the **image compression L_p-metric** is the usual L_p**-metric** on $\mathbb{R}^{n^2}_{\geq 0}$ (the set of gray-scale images seen as $n \times n$ matrices) with p being a solution of the equation $r = \frac{p-1}{2p-1} \cdot e^{\frac{p}{2p-1}}$. So, $p = 1, 2, \infty$ for, respectively, $r = 0, r = \frac{1}{3}e^{\frac{2}{3}} \approx 0.65$, $r \geq \frac{\sqrt{e}}{2} \approx 0.82$. Here r estimates **informative** (i.e., filled with non-zeros) part of the image. It is a quality metric to select a lossy compression scheme.

- The **digital volume metric** (a digital analog of the **Nikodym metric**) on bounded subsets (images) of $\mathbb{Z}^n$) is $vol(A \triangle B)$, where $vol(A) = |A|$ (number of pixels in A), and $A \triangle B$ is the **symmetric difference** of sets A and B.

Hausdorff distance



http://en.wikipedia.org/wiki/User:Rocchini

Consider two binary images, seen as non-empty subsets A and B of a finite metric space (say, a raster of pixels) (X, d).

- Their Baddeley's **p-th order mean Hausdorff distance** is $\left(\frac{1}{|X|}\sum_{x\in X}|d(x,A)-d(x,B)|^p\right)^{\frac{1}{p}}$, where $d(x,A)=\min_{y\in A}d(x,y)$. For $p=\infty$, it is proportional to usual Hausdorff metric.
- Their Dubuisson-Jain's **modified Hausdorff distance** is $\max\left\{\frac{1}{|A|}\sum_{x\in A}d(x,B), \frac{1}{|B|}\sum_{x\in B}d(x,A)\right\}$.
- If $|A|=|B|=m$, $\min_f \max_{x\in A} d(x,f(x))$, where f is any bijective mapping between A and B, is their **bottleneck distance**.

  Variations of above distance are: **minimum weight matching** $\min_f \sum_{x\in A} d(x,f(x))$, **uniform matching** $\min_f(\max_{x\in A} d(x,f(x))\text{-}\min_{x\in A} d(x,f(x))$ and **minimum deviation matching** $\min_f(\max_{x\in A} d(x,f(x))\text{-}\ \frac{1}{|A|}\sum_{x\in A}\ d(x,f(x))$.

Consider two two images, seen as non-empty compact subsets A and B of a metric space (X, d).

- Their **non-linear Hausdorff metric** (or **wave distance**) is the **Hausdorff distance** $d_{Haus}(A \cap B, (A \cup B)^*)$, where $(A \cup B)^*$ is the subset of $A \cup B$ which forms a closed contiguous region with $A \cap B$, and the distances between points are allowed to be measured only along paths wholly in $A \cup B$.

- Their **Hausdorff distance up to** G, for given group $(G, \cdot, id)$ acting on the Euclidean space $\mathbb{E}^n$, is $\min_{g \in G} d_{Haus}(A, g(B))$. Usually, G is the group of all isometries or all translations of $\mathbb{E}^n$.

- Their **hyperbolic Hausdorff distance** is the **Hausdorff metric** between $MAT(A)$ and $MAT(B\text{MAT(A)})$ of (X, d_{hyp}), where the **hyperbolic distance** $d_{hyp}(x, y)$ is $\max\{0, d_E(x', y') - (r_y - r_x)\}$ for elements $x = (x', r_x)$ and $y = (y', r_y)$ of X.

  Here $MAT(C)$ denotes, for any compact $C \subset \mathbb{R}^n$, its **Blum's medial axis transform**, i.e., the subset of $X = \mathbb{R}^n \times \mathbb{R}_{\geq 0}$ of all pairs $x = (x', r_x)$ of the centers x' and the radii r_x of the maximal inscribed (in C) l_2-balls, in terms of the Euclidean distance d_E in $\mathbb{R}^n$.

- Let (X, d) be a metric space, and let $M \subset X$.

  The set $MA(X) = \{x \in X : |\{m \in M : d(x, m) = d(x, M)\}| \geq 2\}$ is the **medial axis** of X. It consists of all points of boundaries of **Voronoi regions** (**zones of influence**) of points of M.

  The **cut locus** of X is its closure. The **medial axis transform** is the point-weighted set $MA(X)$ (the restriction of the **distance transform** on $MA(X)$) with $d(x, M)$ being the weight of $x \in X$.

  If (as usual in applications) $X \subset \mathbb{R}^n$ and M is its boundary, then the **skeleton** $Skel(X)$ of X is the set of the centers of all d-balls inscribed in X and not belonging to any other such ball; so, $Skel(X)$=$MA(X)$.

  For $2D$ binary images X, the skeleton is a curve, a single-pixel thin one, in digital case.

  The **exoskeleton** of X is the skeleton of the complement of X, i.e., of the background of the image for which X is the foreground.

- Given a metric space (X, d) ($X = \mathbb{Z}^2$ or $\mathbb{R}^2$) and a binary image $M \subset X$, the **distance transform** (or **distance field**, **distance map**) is a function $f_M : X \to \mathbb{R}_{\geq 0}$, where $f_M(x) = d(x, M) = \inf_{u \in M} d(x, u)$. So, it can be seen as a gray-scale image where pixel gray-level is labeled by its distance to the nearest pixel of the background.

  The **Voronoi surface** of M is $\{(x, d(x, M)) : x \in X = \mathbb{R}^2\}$.

- Let see two digital images as binary $m \times n$ matrices $x = ((x_{ij}))$ and $y = ((y_{ij}))$, where a pixel x_{ij} is black or white if it is 1 or 0, resp.

  For each pixel x_{ij}, the **fringe distance map** to the nearest pixel of opposite color $D_{BW}(x_{ij})$ is the number of **fringes** expanded from (i, j) (where each fringe consists of pixels that are equi-distant of (i, j)) until the first fringe with a pixel of opposite color is reached. Then $\sum_{1 \leq i \leq m} \sum_{1 \leq j \leq n} |x_{ij} - y_{ij}|(D_{BW}(x_{ij}) + D_{BW}(y_{ij}))$ is **pixel distance**.

- In any metric space (X, d), the **point-set distance** $d(x, M)$ between $x \in X$ and $M \subset X$ is $\inf_{y \in M} d(x, y)$.

  The function $f_M(x) = d(x, M)$ is a (general) **distance map**.

- The **set-set distance** between two subsets $A, B \subset X$ is $\inf_{x \in A,} d(x, B)$. The **Hausdorff metric** is $\max\{d_{dHaus}(A, B), d_{dHaus}(B, A)\}$, where $d_{dHaus}(A, B) = \max_{x \in A} \min_{y \in B} d(x, y)$ (for compact subsets $A, B \subset X$).

- If the boundary $B(M)$ of the set M is defined, then

  the **signed distance function** g_M is defined as $-\inf_{u \in B(M)} d(x, u)$ for $x \in M$ and $\inf_{u \in B(M)} d(x, u)$, otherwise.

  If M is a (closed and orientable) manifold in $\mathbb{R}^n$, then g_M is the solution of the **eikonal equation** $|\nabla g| = 1$ for its **gradient** ∇.

- The shape can be represented by a parameterized simple plane curve. Let $X = X(x(t))$, $Y = Y(y(t))$ be two parameterized curves, where $x(t)$, $y(t)$ are continuous on $[0, 1]$ and $x(0) = y(0) = 0$, $x(1) = y(1) = 1$. The most used **parameterized curves distance** is the minimum, over all monotone increasing $x(t)$, $y(t)$, of $\max_t d_E(X(x(t)), Y(y(t)))$. It is Euclidean case of the **dogkeeper distance** which is, in turn, the **Fréchet metric** for the case of curves.

- Consider a digital representation of curves. Fix $r \geq 1$ and let $A = \{a_1, \ldots, a_m\}$, $B = \{b_1, \ldots, b_n\}$ be finite ordered sets of consecutive points on two closed curves. For any order-preserving correspondence f between all points of A and B, the **stretch** $s(a_i, b_j)$ *of* $(a_i, f(a_i) = b_j)$ is r if either $f(a_{i-1}) = b_j$ or $f(a_i) = b_{j-1}$, or zero, otherwise.

  The **elastic matching distance** is $\min_f \sum(s(a_i, b_j) + d(a_i, b_j))$, where $d(a_i, b_j)$ is the difference between the tangent angles of a_i and b_j. It is a **near-metric** for some r: all $d(x, y) \leq C(d(x, z) + d(z, y))$ for $C \geq 1$.

- For a plane polygon P, its **turning function** $T_P(s)$ is the angle between the counterclockwise tangent and the x-axis as the function of the arc length s. This function increases with each left hand turn and decreases with right hand turns.

  Given two polygons of equal perimeters, their **turning function distance** is the L_p**-metric** between their turning functions.

- For a plane graph $G = (V, E)$ and a **measuring function** f on its vertex-set V (say, the distance from $v \in V$ to the center of mass of V), the **size function** $S_G(x, y)$ on the points $(x, y) \in \mathbb{R}^2$ is the number of connected components of the restriction of G on vertices $\{v \in V : f(v) \leq y\}$ containing a point v' with $f(v') \leq x$.

  Given two plane graphs with vertex-sets belonging to a raster $R \subset \mathbb{Z}^2$, their Uras-Verri's **size function distance** is the normalized l_1-metric between their size functions over raster pixels.

- The **time series video distances** are objective wavelet-based spatial-temporal **video quality metrics**.

  A video stream x is processed into time series $x(t)$ (seen as a curve on coordinate plane) which then (piecewise linearly) approximated by a set of n contiguous line segments that can be defined by $n+1$ endpoints (x_i, x'_i), $0 \leq i \leq n$, on coordinate plane.

  Wolf-Pinson's distances between video streams x and y are:

  1. $Shape(x,y) = \sum_{i=0}^{n-1} |(x'_{i+1} - x'_i) - (y'_{i+1} - y'_i)|$;
  2. $\text{Offset}(x,y) = \sum_{i=0}^{n-1} |\frac{x'_{i+1}+x'_i}{2} - \frac{y'_{i+1}+y'_i}{2}|$.

Representation of distance in Painting

In Western Visual Arts, the **distance** is the part of a picture representing objects which are the farthest away, such as a landscape; it is the illusion of 3D depth on a flat picture plane. The **middle distance** is the central part of a scene between the foreground and the background (implied horizon).

Perspective projection draw distant objects as smaller to provide additional realism by matching the decrease of their visual angle.

A **vanishing point** is a point at which parallel lines receding from an observer seem to converge. **Linear perspective** is a drawing with $1-3$ such points; usually, they placed on horizon and equipartition it.

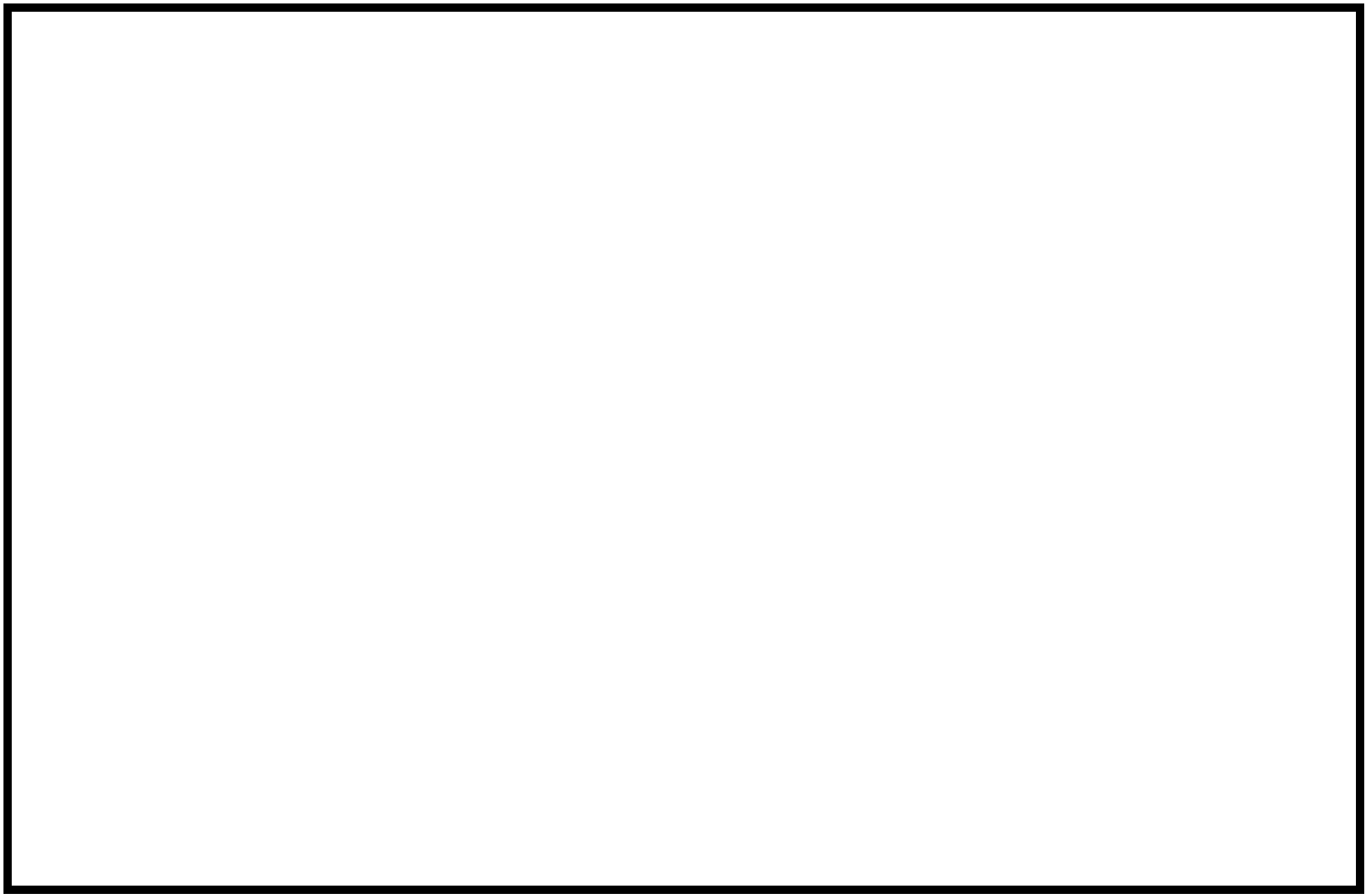
In a **curvilinear perspective**, there are ≥ 4 vanishing points; usually, they mapped into and equipartition a **distance circle**. **0-point perspective** occurs if the vanishing points are placed outside the painting, or if the scene (say, a mountain range) does not contain any parallel lines.

Axonometric projection is parallel projection which is *orthographic* (i.e., the projection rays are perpendicular to the projection plane) and such that the object is rotated along one or more of its axes relative to this plane. The main case of it, used in engineering drawing, is **isometric projection** in which the angles between three projection axes are the same, or $\frac{2\pi}{3}$.

In **Chinese painting**, the **high-distance**, **deep-distance** or **level-distance** views correspond to picture planes dominated, respectively, by vertical, horizontal elements or their combination.

Instead of the perspective projection of a "subject", assuming a fixed position by a viewer, Chinese classic hand scrolls (up to 10 m in length) used axonometric one. It permitted to move along a continuous/seamless visual scenario and to view elements from different angles.

It was less faithful to appearance and allowed to present only 3 (instead of 5) of 6 surfaces of a normal interior. But in Chinese painting, the focus is rather on symbolic and expressionist representation.



DICTIONARY OF DISTANCES

The concept of distance is one of the basic ones in human experience; this is the first book treating it in full generality.

Distance metrics have become an essential tool in many areas of Mathematics and its applications. However, a lot of information on distances is too scattered and hardly accessible for non-experts. In view of the growing need (especially in Information Retrieval with respect to Image, Audio, Internet and Biology) for an accessible interdisciplinary source on distances, we have expanded our private collection into this Dictionary.

It aims to be a thought-provoking archive: besides distances, many distance-related notions and paradigms are collected also, in ready-to-use fashion.

In a time when over-specialization and the use of terminology isolates researchers, this Dictionary tries to be "centripetal" and "ecumenical", providing some access and altitude of vision but without taking the route of scientific vulgarization.

The Dictionary is divided into 28 chapters grouped into 7 Parts. Parts II, III and IV, V require some culture in, respectively, pure and applied Mathematics. Part VII can be read by a layman.

The chapters are thematic lists which can be read independently. When necessary, a chapter or a section starts with a short introduction. Each chapter consists of items ordered in a way that hints of connections between them. All item titles and key terms can be traced via the Index.

Many nice curiosities appear in this "Who is Who" of distances. Also distances having physical meaning show up; they range from 1.6×10^{-35} m (Planck length) to 7.4×10^{26} m (estimated size of the observable Universe).

The target audience consists of all researchers working on some measuring schemes, of students and the general public interested in science.

ELSEVIER books.elsevier.com

ISBN 0 444 52087 2
9 780444 520876

DEZA
DEZA

DICTIONARY OF DISTANCES

ELSEVIER

Dictionary of DISTANCES

Elena Deza and
Michel Marie Deza

ELSEVIER