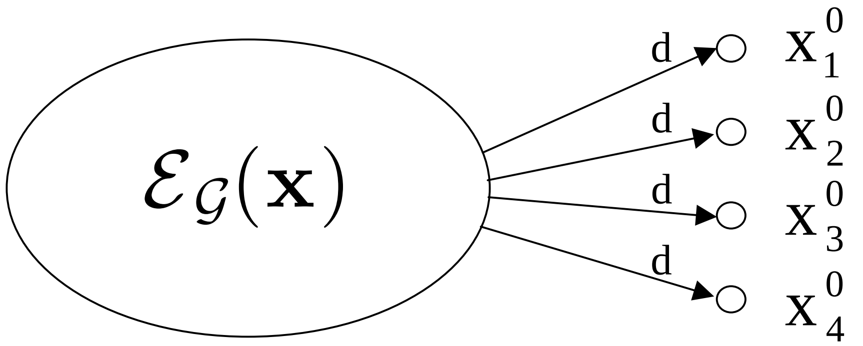
Classification through direct comparison (Matching)

A trivial solution for classification is direct matching of the unknown pattern in all its appearances of its equivalence class (pattern matching) corresponding to metrics d (measure for class-affiliation) with the prototypes of a class (e.g.: jigsaw puzzle)

equivalence class: $$\mathcal{E}_{\mathcal{G}}(\mathbf{x}) := \{g_i(\mathbf{x}) \mid \forall g_i \in \mathcal{G}\}$$



Introducing a metric d(**x**,**y**) for comparing in unitary or inner product spaces (or more general in Hilbert-spaces):
(the inner or scalar product induces a metric)

$$d(\mathcal{E}_{\mathcal{G}}(\mathbf{x}),\mathbf{x}_j^0) = \left\| \mathcal{E}_{\mathcal{G}}(\mathbf{x}) - \mathbf{x}_j^0 \right\| = < \mathcal{E}_{\mathcal{G}}(\mathbf{x}) - \mathbf{x}_j^0, \mathcal{E}_{\mathcal{G}}(\mathbf{x}) - \mathbf{x}_j^0 >^{1/2}$$

$$= \sqrt{\left\| \mathcal{E}_{\mathcal{G}}(\mathbf{x}) \right\|^2 + \left\| \mathbf{x}_j^0 \right\|^2 - 2 < \mathcal{E}_{\mathcal{G}}(\mathbf{x}), \mathbf{x}_j^0 >} = d_j(\mathbf{x}(\mathbf{p}), \mathbf{x}_j^0) \quad \text{for } \forall \mathbf{p}$$

The smallest distance yields a greatest inner product for normed vectors (metric for “similarity”).

Scalar products for continuous and digitized Patterns:

continuous patterns:

$$\text{1D:} \quad <\mathbf{x},\mathbf{y}> = \int x(t)y(t)dt$$

$$\text{2D:} \quad <\mathbf{X},\mathbf{Y}> = \iint X(t_1,t_2)Y(t_1,t_2)dt_1dt_2$$

digitized patterns:

$$\text{1D:} \quad <\mathbf{x},\mathbf{y}> = \sum x_i y_i$$

$$\text{2D:} \quad <\mathbf{X},\mathbf{Y}> = \sum\sum X_{i,j}Y_{i,j}$$

Calculating the Scalar Product over the complete Parameter space {**p**} of the motion group corresponds with calculating the cross-correlation function $R_{x,y}$

$$<\mathcal{E}_{\mathcal{G}}(\mathbf{x}), \mathbf{x}_j^0> = <\mathbf{x}(\mathbf{p}), \mathbf{x}_j^0> = R_{\mathbf{x}(\mathbf{p}), \mathbf{x}_j^0} \quad \text{für} \ \forall \mathbf{p}$$

e.g. translating pictures of dimension N×N has complexity proportional to:

$$\underbrace{N^2}_{\text{cyclic permut.}} \cdot \underbrace{N^2}_{\text{scalar product}} = N^4$$

The correlation can be calculated more efficiently with fast Fourier transform, with complexity $O(N^2 \, ld \, N)$ according to (see DBV-I):

$$R_{x,y} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{x}) \cdot \mathcal{F}^*(\mathbf{y}))$$

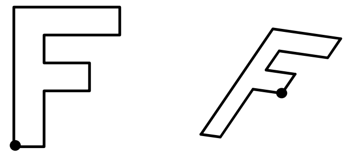
Calculating the scalar product over the complete parameter space $\{\mathbf{p}\}$ of the motion group has for a signal or picture of dimension N in general this compexity :

$$O(N \cdot N^{\dim(\mathbf{p})})$$

cost for scalar product

It is assumed, that every axis of the parameter space is resolved with the image dimesion, e.g. N tranlations.

Affine invariant recognition with correlation in all parameters (matched filter)



Given: boundary with N points

➔ $O(N \times N^{\dim(\mathbf{p})}) = O(N^8)$ ops with $\dim(\mathbf{p})$: degrees of freedom

e.g. $N = 10^3 \Rightarrow \sim 10^{24}$ ops

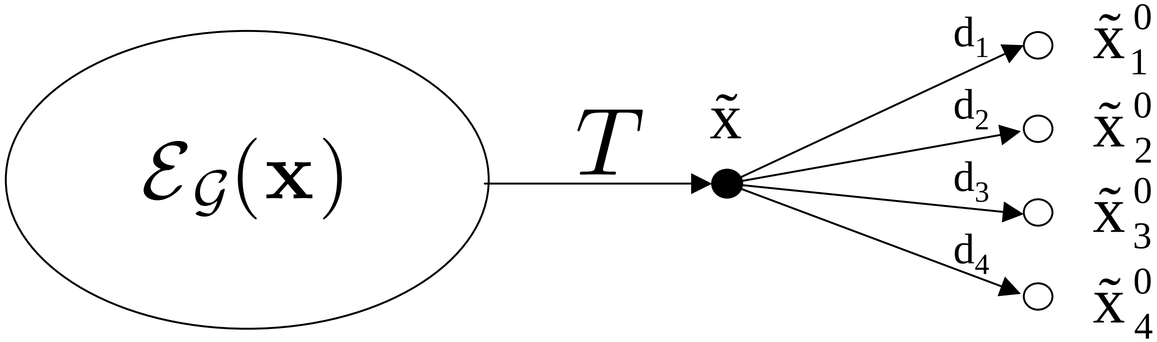
Assumption: 100 MIPS machine (multiply-add)

$$\Rightarrow \sim \left(\frac{10^{24}\ \text{ops}}{10^8\ \text{ops/sec}}\right) = 10^{16}\ \text{sec} = 3.2 \times 10^8\ \text{years}$$

age of cosmos: 5×10^9 years

Classification with invariant features:

Extraction of location invariant features, afterwards classification in the feature space:



$$\tilde{\mathbf{x}} = T(\mathbf{x})$$

Necessary condition for invariance of a mapping T :

1) Necessary condition for invariance with respect to group action $\mathcal{G}$**:**

$$\mathbf{x}_1 \overset{\mathcal{G}}{\sim} \mathbf{x}_2 \Rightarrow T(\mathbf{x}_1) = \tilde{\mathbf{x}}_1 = \tilde{\mathbf{x}}_2 = T(\mathbf{x}_2)$$

All elements of an equivalence class are mapped on a point in the feature space. This can be expressed as follows:

$$T(g_i \mathbf{x}) = T(\mathbf{x}) \quad \text{für} \quad \forall g_i \in \mathcal{G}$$

Definition of invariance with respect to time invariant systems

The idea of invariance as introduced above must not be confused with the defintion of time invariant systems. This requires the systems response to a staggered excitation with unaltered ***form*** to be also staggered. This holds both for excitation with a δ-impulse and any excitation x, i.e.:

$$\delta(t) \stackrel{T}{\Longrightarrow} h(t)$$

$$\delta(t-t_0) \stackrel{T}{\Longrightarrow} h(t-t_0)$$

resp.:

$$x(t) \stackrel{T}{\Longrightarrow} y(t)$$

$$x(t-t_0) \stackrel{T}{\Longrightarrow} y(t-t_0)$$

Linear Operator (Homomorphism)

For time invariant systems the homomorphism is a linear operator, so the following holds:

$$T(g_i(x(t))) = g_i(T(x(t))) = g_i(y(t))$$

with: $$g_i(x(t)) = x(t - t_i)$$

And thus: the operations g and T can be interchanged!

Completeness of an invariant mapping T :

2) The ***completeness*** is sufficient for assigning two objects to the same equivalence class based on the equality of two feature vectors (reversing the sine qua non):

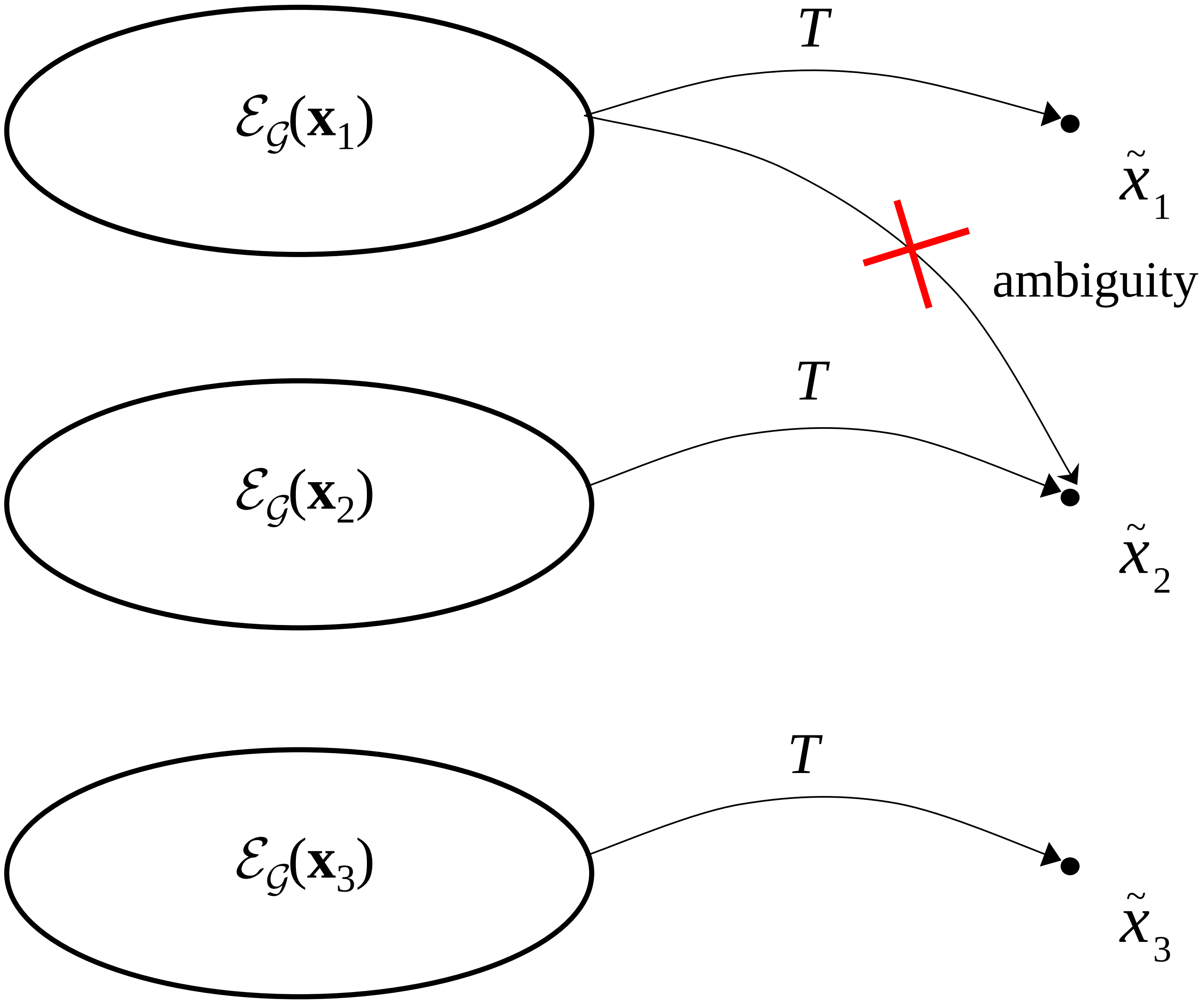
$$T(\mathbf{x}_1) = T(\mathbf{x}_2) \Rightarrow \mathbf{x}_1 \overset{\mathcal{G}}{\sim} \mathbf{x}_2$$

This means that there are no ambiguities for assigning feature vectors to equivalence classes.

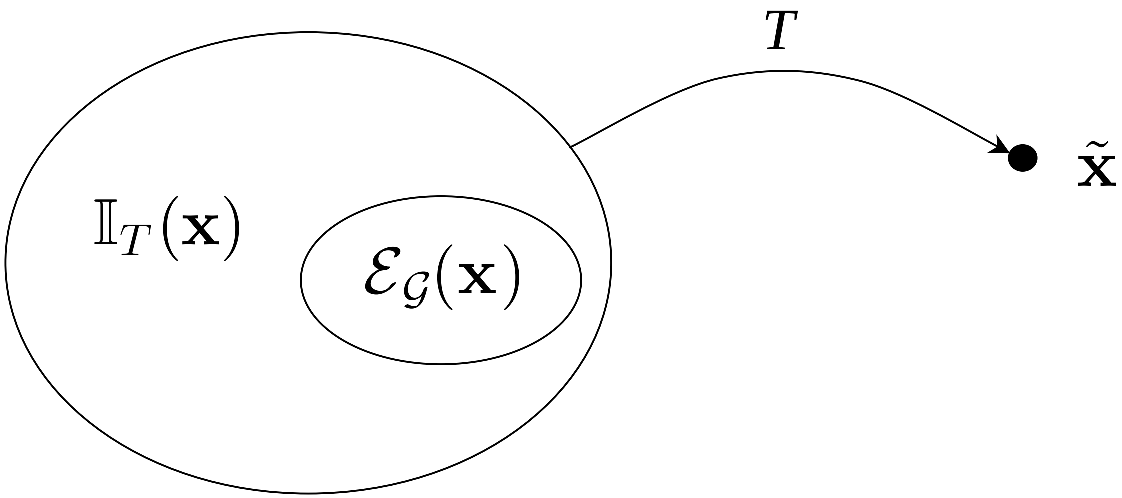
Weaker constraint: ***Separability*** *(completeness for a subset)*

e.g. The set of all letters for sign recognition

Properties of the transformation T:



Map properties: Measure for degree of completeness



1. Definition of the invariants of an arbitrary contractive map *T*:

$$\mathbb{I}_T(\mathbf{x}) := \{\mathbf{x_i} \mid T(\mathbf{x_i}) = T(\mathbf{x})\}$$

Map properties

2. From the invariance constraint follows:

$$\mathcal{E}_{\mathcal{G}}(\mathbf{x}) \subseteq \mathbb{I}_T(\mathbf{x}) \qquad \text{must be a subset!}$$

3. From completeness follows:

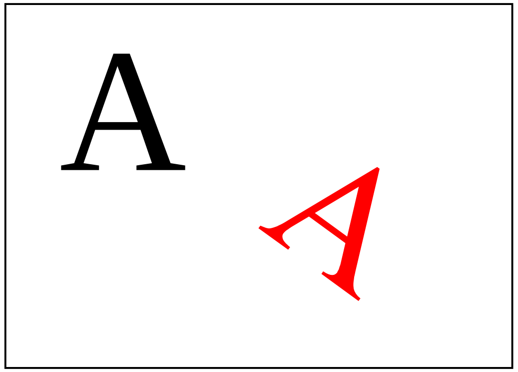
$$\mathcal{E}_{\mathcal{G}}(\mathbf{x}) = \mathbb{I}_T(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathbb{X}$$

For the weaker condition "separability" this holds only for a subset $\mathbb{X}_0$.

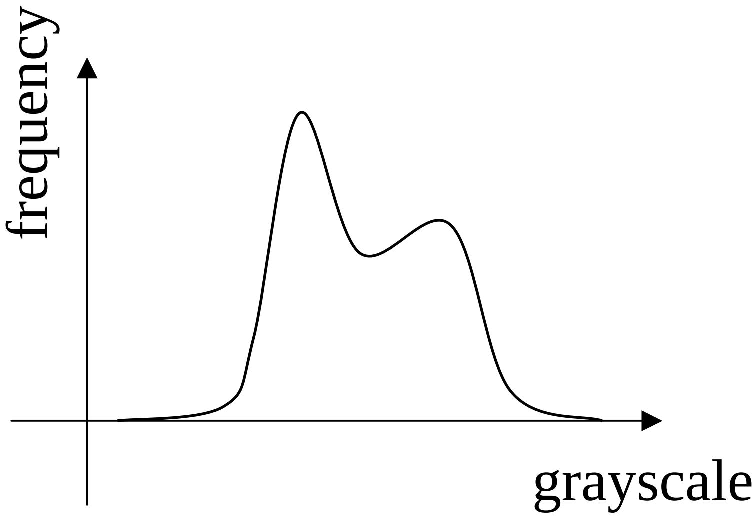
Examples for location invariant features

1. The mean value of images is invariant with respect to Euclidian motions (translation and rotation). But: features are not complete.

$$\bar{x} = \frac{1}{N} \sum \sum x_{i,j}$$

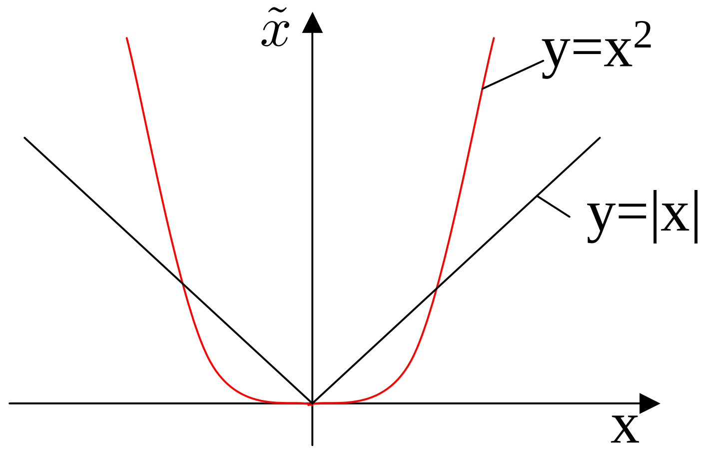


2. The histogram or frequency distribution is also invariant. Also, the neccessary condition is fulfilled, but the features are ambigous, because all pixels can be permuted without changing the histogram. The histogram is complete for the symmetric group (all N! permutations).



Sentence: Only non-linear mappings can form invariants from limited equivalence classes.

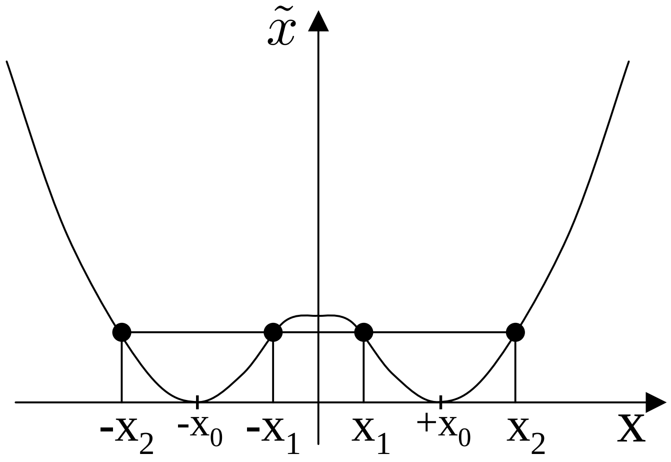
(Simple) Examples for invariant maps with non-linear functions



The equivalence class is formed based on absolute values, e.g.:

$$(-3) \overset{|\bullet|}{\sim} (+3)$$

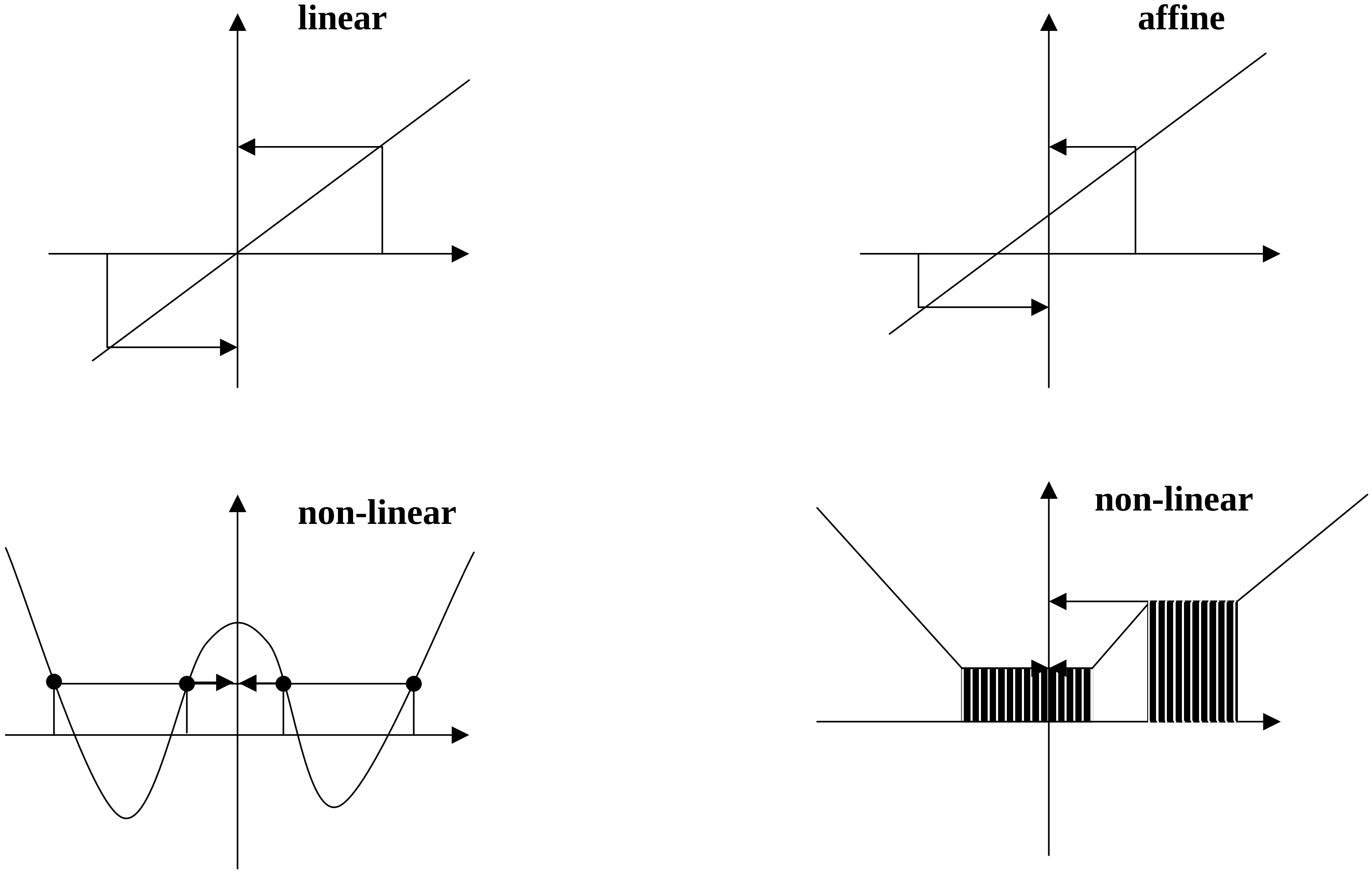
complete invariants!



$$y = a_0 + a_2 x^2 + a_4 x^4$$

Holds ***necessary***, but not sufficient conditions; ambiguities. But e.g. ***separability*** for $|x| < x_0$

Non-linear maps for generating invariants



Further requirements to invariant maps

- continuity of the map
- preserve clusters
- What happens when systematical noise occurs? (brightness and contrast jitter)

$$\boxed{\mathbf{x} = k\,\mathbf{x}^0 + a\,\mathbf{u}} \qquad \text{with}: \quad \mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Linear systems can distinguish between “frequencies” or spectral components,

non-linear systems can distinguish between “equivalence classes”.

- **Lemma:**
  Given a linear map *S*. Two signals x_1,x_2 are equivalent to each other, if and only if the difference x_1-x_2 is an element of the null space *N(S)*.
- ....
- It is impossible to represent/describe nontrivial compact equivalence classes with linear maps!