Independent Mechanism Analysis: Context and Methods

Shashank Shetty Kalavara

Supervisor: Dr. Nico Scherf & Nikola Milosevic

Neural Data Science group at MPI CBS Leipzig

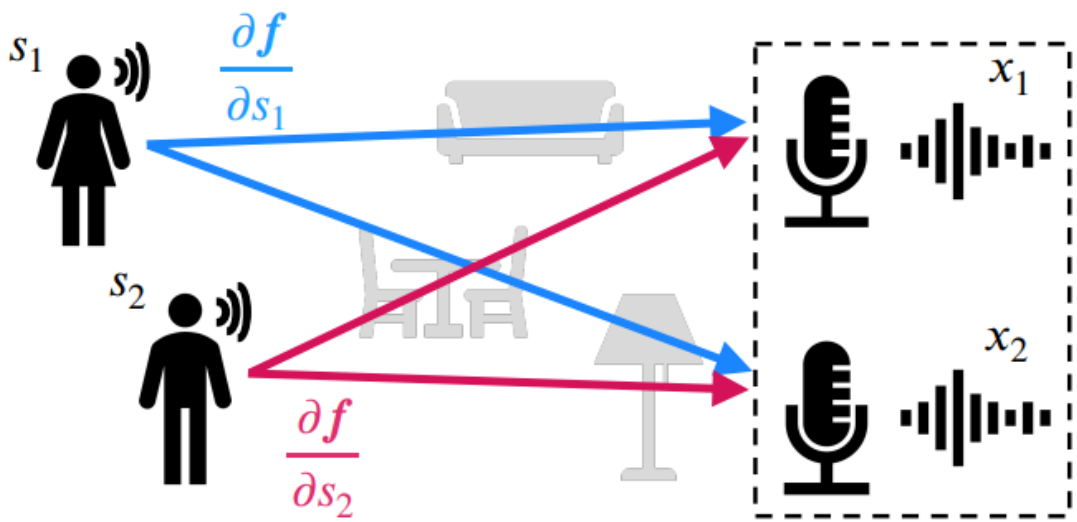
May 22, 2023

MAX-PLANCK-GESELLSCHAFT

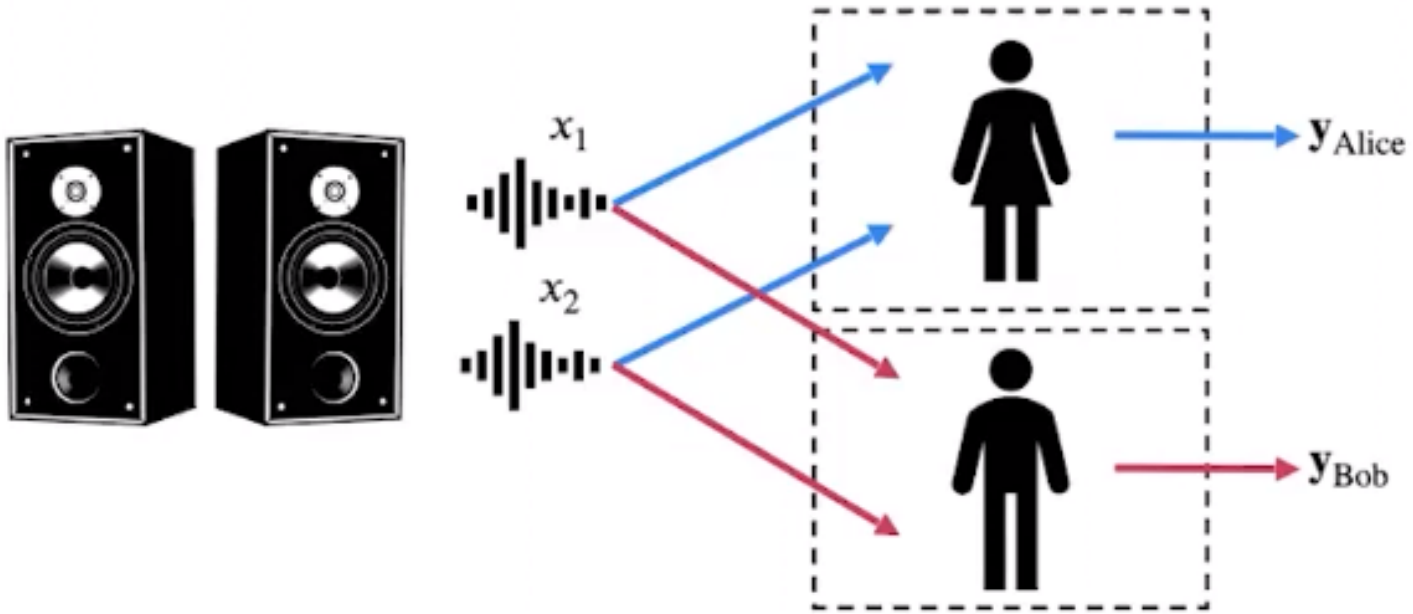
Contents

Blind Source Separation

Separate a set of source signals from their observed mixtures without relying on prior knowledge about the sources or the mixing process. The term "blind" refers to the lack of information about the mixing process and the nature of the source signals.



The Independent-Listeners Problem

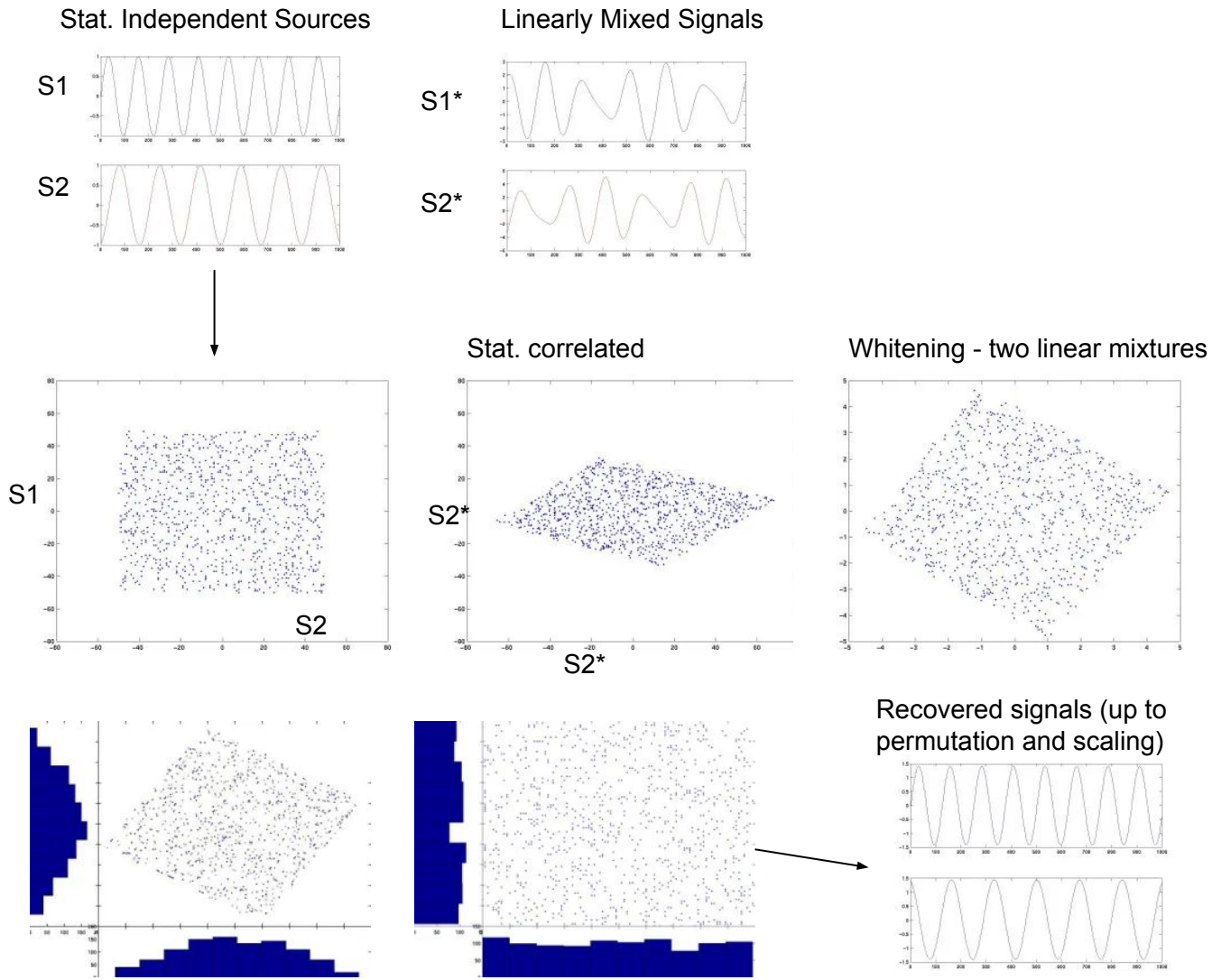


1

[1] Roeder, Geoffrey, Luke Metz, and Diederik P. Kingma. 2020. "On Linear Identifiability of Learned Representations."

Independent Component Analysis (ICA)

2 dim. example - minimizes the Gaussianity - rotation



[2] Laurenz Wiskott

Identifiablility

Definition: Models Statistical Identifiability

A model class is said to be statistically Identifiable if
$\forall s, s' \in I : \; p_s(x) = p_{s'}(x) \;\; \forall x \;\; \rightarrow \;\; x = x'$

Identifiablility in essence means a methods like say ICA can do blind source separation up to tolerable ambiguities.

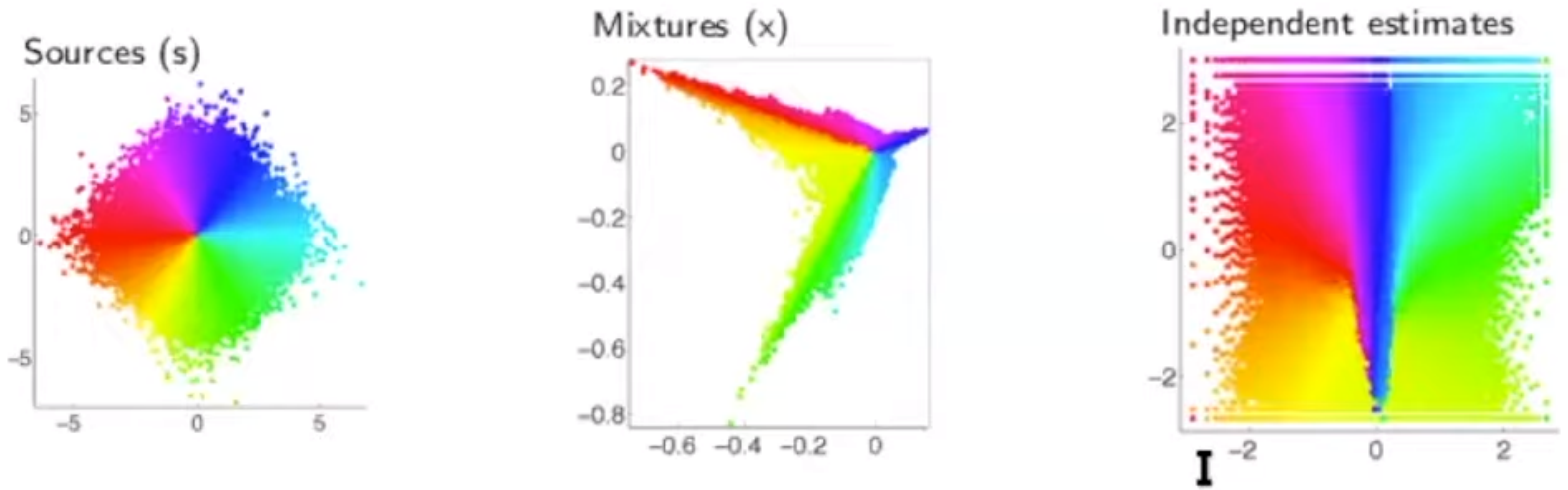
Given the set of rules in a method of decomposition - is it possible to recover - **a set of unique solutions** to the source and mixing. More often than not methods are non-identifiable.

Tolerable/Unavoidable ambiguities in linear ICA:

- Scaling
- Permutation

Nonlinear ICA Not Working

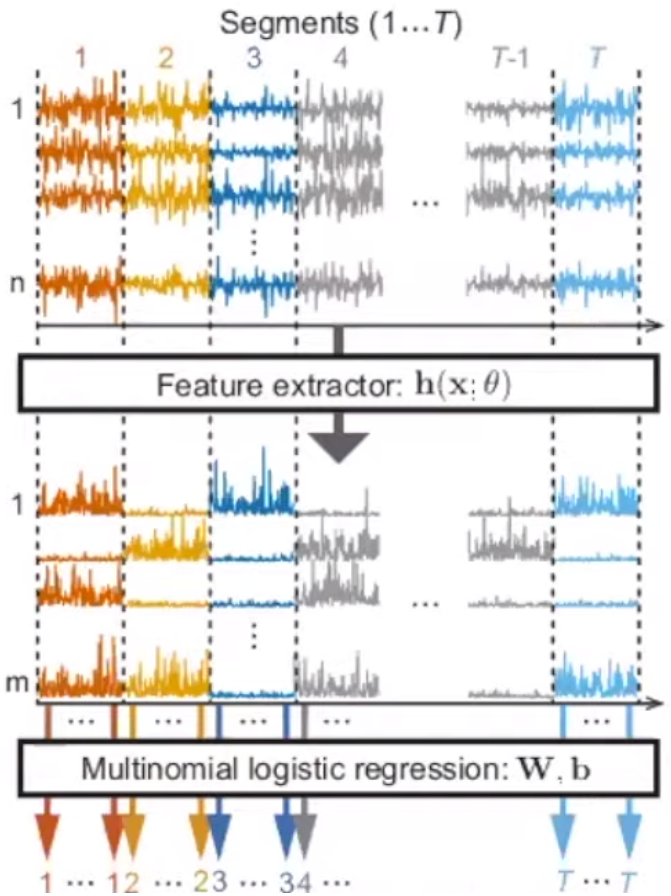
Example of a failed Nonlinear ICA:



Most Nonlinear ICA's are not identifiable, hence they tend to recover spurious solutions to source and mixing terms.

Nonlinear ICA Working

Example of a somewhat successful Nonlinear ICA. Time-Contrastive Learning (TCL) method.



Limitations: Does not work i.i.d data, loosely Identifiable up to source squaring. [3]

[3] TCL

Identifiability in Nonlinear ICA

Non-linear ICA methods are generally not statistically Identifiable, so they made a looser definition for identifiability in terms of equivalence relation.

Definition: Equivalence relation

An equivalence relation on set A is a binary relation, equivalent or false, which should satisfy, for $\forall a, b, c \in A$

- Reflexivity $a\ a$
- Symmetry $a\ b \rightarrow b\ a$
- Transitivity $(a\ b) \wedge (b\ c) \rightarrow a\ c$

An Equivalence relation is less strict than the statistical identifuability. As equivalence relation on set A imposes partition into disjoint subsets, each corresponding to an equivalence class.

Which leads to the following: **Tolerable/Unavoidable ambiguities in Nonlinear ICA**:

- Scaling
- Permutation
- Non linearly Transformed sources
  $h(s_1)$,$h(s_2)$...$h(s_n)$, where 'h' is a nonlinear function for 'n' sources

Identifiability in IMA with Auxiliary Variable

Identifiability in IMA in essence is very similar to the case as in earlier slide but with an additional caveat, of Auxiliary variable restriction **i.e cannot do BSS for i.i.d's.**

Introduction of an auxiliary variable **U** allows to make the sources conditionally independent.

$$i.i.d\left[s \sim P_{s|u}\right], \qquad P_{s|u} = \prod_{i=1}^{n} \left(P_{s_i|u}\left(s_i|u\right)\right)$$

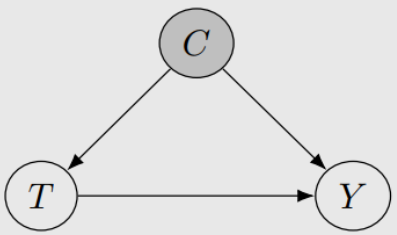
With suitable assumptions identifiability can be achieved without restricting the mixing function, sometime up to "Equivalence class" Blind source separation.

Simpson's Paradox

Application to the COVID-27 example

$$\mathbb{E}[Y|\text{do}(T=t)] = \mathbb{E}_C\mathbb{E}[Y|t,C] = \sum_c \mathbb{E}[Y|t,c]P(c)$$

Causal Graph



Treatment	Mild	Severe	Total	Causal
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)	19.4%
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)	**12.9%**
	$\mathbb{E}[Y\|t,C=0]$	$\mathbb{E}[Y\|t,C=1]$	$\mathbb{E}[Y\|t]$	$\mathbb{E}[Y\|\text{do}(t)]$

Condition

$$\frac{1450}{2050}(0.15) + \frac{600}{2050}(0.30) \approx 0.194$$

$$\frac{1450}{2050}(0.10) + \frac{600}{2050}(0.20) \approx 0.129$$

4

[4] Brady Neal

The Principle of Independent Mechanisms

Effect of Intervention

Mapping relation between **Altitude** and **Temperature**:

$$P(a, t) = P(a|t)P(t)$$

$P(a|t)$:T $\rightarrow$ A

$$P(a, t) = P(t|a)P(a)$$

$P(t|a)$:A $\rightarrow$ T

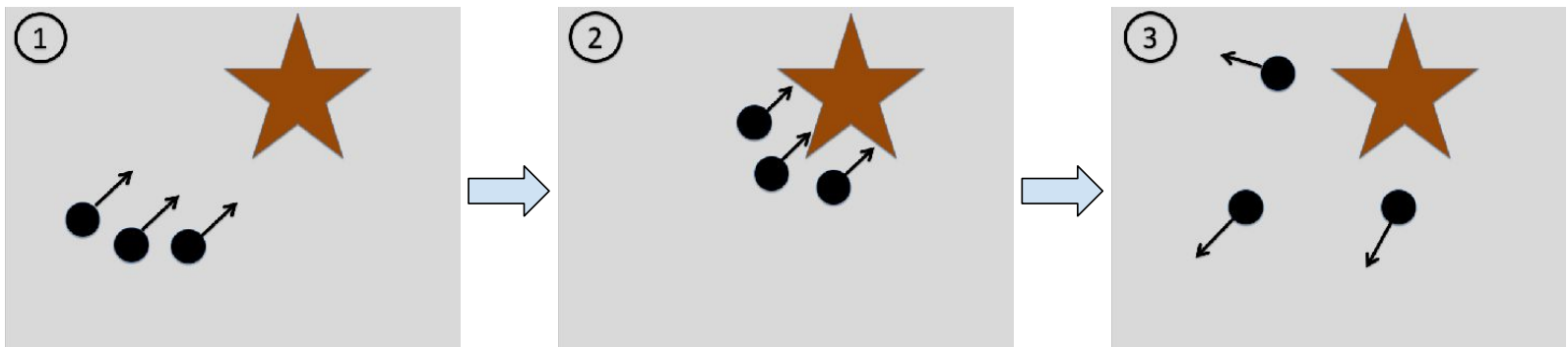
which of the two structures is the causal one?

Intervention establishes the causal relation to be $P(t|a)$:A $\rightarrow$ T [5]

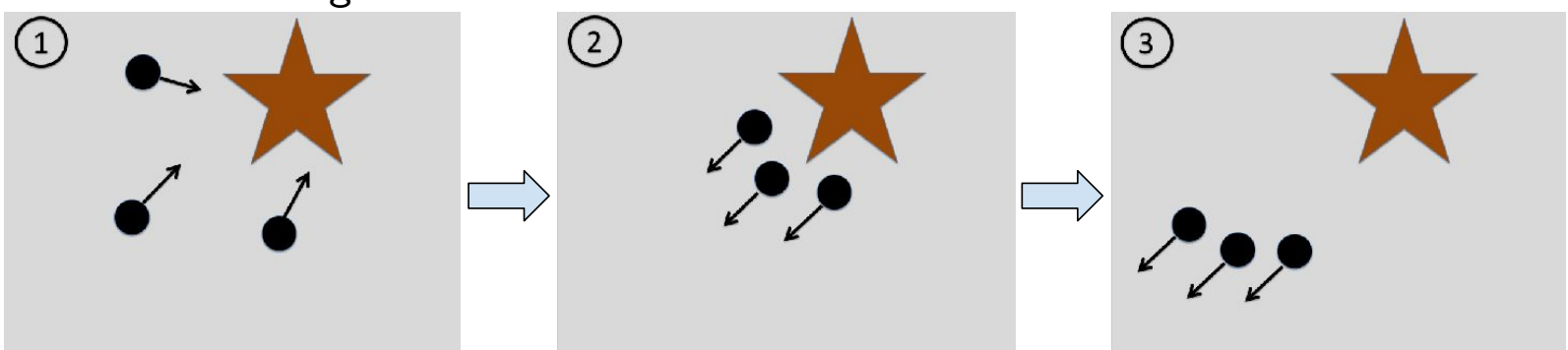
[5] Elements of Causal Inference, 2.1

ICM and the Thermodynamic Arrow of Time

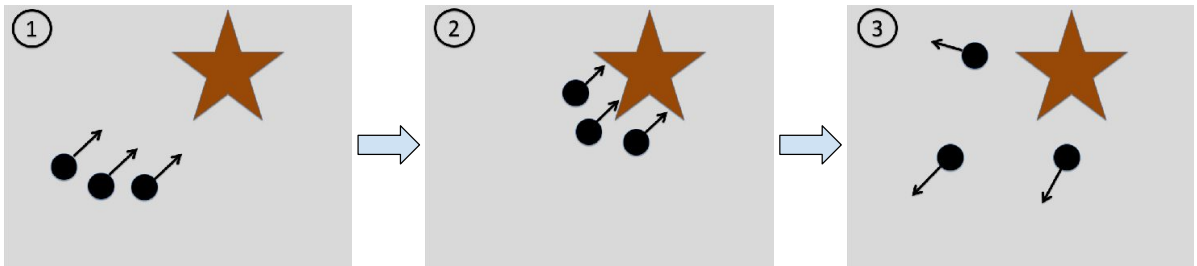
Initial state and dynamical law: If s is the initial state of a physical system and M a map describing the effect of applying the system dynamics for some fixed time, then s and M are independent. Here, we assume that the initial state, by definition, is a state that has not interacted with the dynamics before.



Reverse scattering



Independence in Cause Effect Relations



The initial state of particle or input signal contains no information about the object/ mixing-Mechanism, and vice versa.

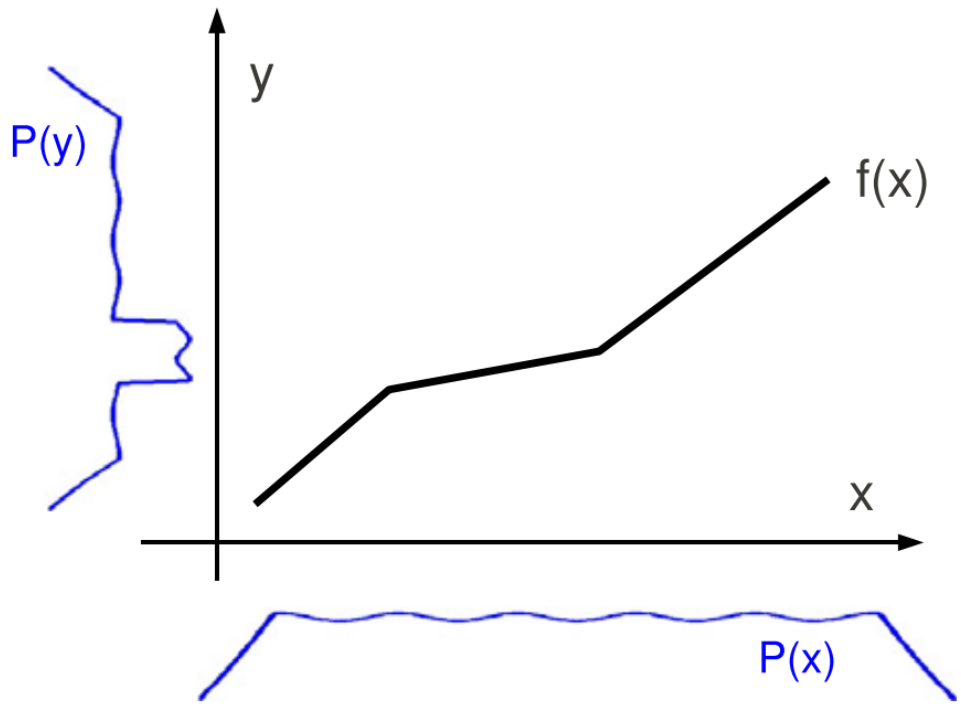
Hence **input** *and the* **mixing mechanism** *are independent from each other.*

Independent Causal Mechanisms: The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.

Inferring Causal Directions with Deterministic Mapping

Information geometric causal inference (IGCI), example in 1-Dimention

D. Janzing et al. / Artificial Intelligence 182–183 (2012) 1–31



If the structure of the density of P_X is not correlated with the slope of f, then flat regions of f induce peaks of P_Y. The causal hypothesis Y $\rightarrow$ X is thus implausible because the causal mechanism f^1 appears to be adjusted to the "input" distribution P_Y.

6

[6] Information-geometric approach to inferring causal directions

Source Separation with ICM

Classical ICM is not useful got Blind Source separation. As it will only impose Independence between all the sources or the cause, and the mixing function or the mechanism.

Formalising ICM with the Information Geometric Causal Inference (IGCI) since IGCI also assumes deterministic mapping between cause and effect.

Formal relation

$$\int log\,|J_f\,(s)|\,p\,(s)\,ds = \int log\,|J_f\,(s)|\,ds \cdot \int p\,(s)\,ds$$

$$= \int log\,|J_f\,(s)|\,ds$$

cause: s and its distribution p(s)
mechanism: f and its Jacobian J_f
effect: x = f(s)

—

Since ICA accomplishes source and mixing separation, it can be expanded to have additional separation between the source terms themselves, accomplishing decomposition!

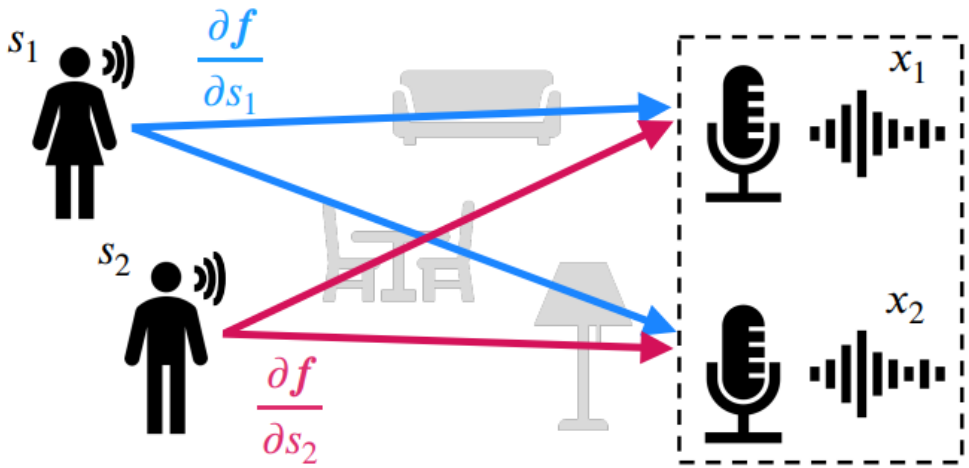
Independent Mechanism Analysis

7

[7] Independent mechanism analysis, a new concept?

Source Separation with ICM

The additional **separation between source** can be achieved by an orthogonality condition:



This is done by imposing Independence on the partial derivatives $\frac{\partial}{\partial s_i}(f)$, where $s = \sum_{i=1}^{n} s_i$ with distribution $p(s) = \sum_{i=1}^{n} \frac{\partial}{\partial s_i}(f)$

Hence the Mechanism by which each source s_i influences the observed distribution is "Independent."

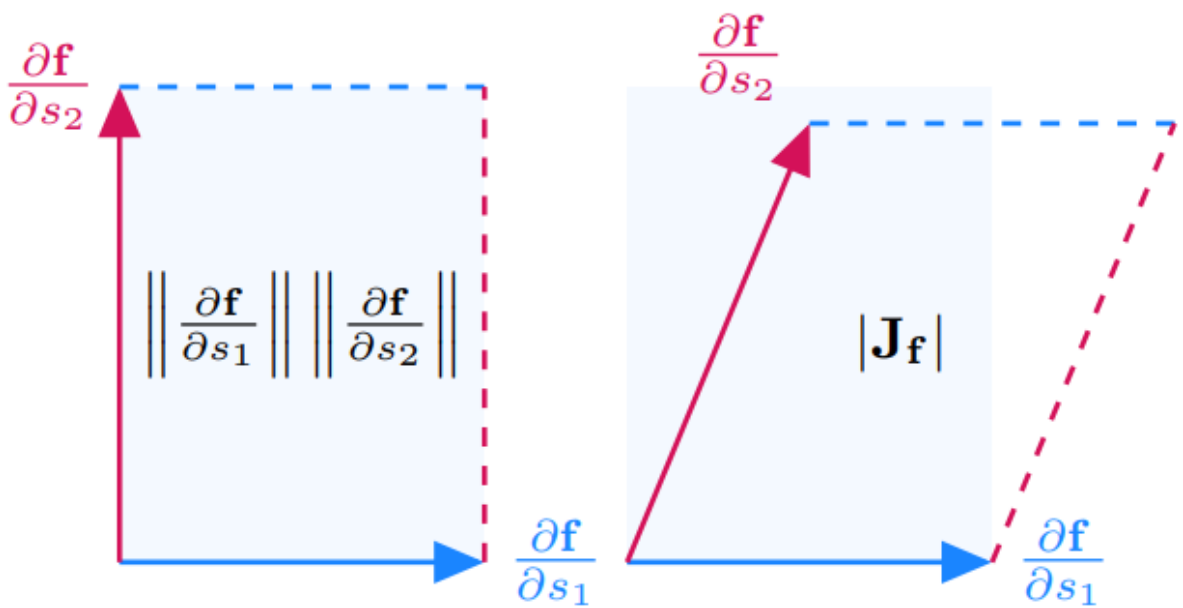
The Principal of Independent Mechanism Analysis

Principal of IMA

The influence of the source terms s_i on the observed distribution is disentangled with the following relation imposing independence:

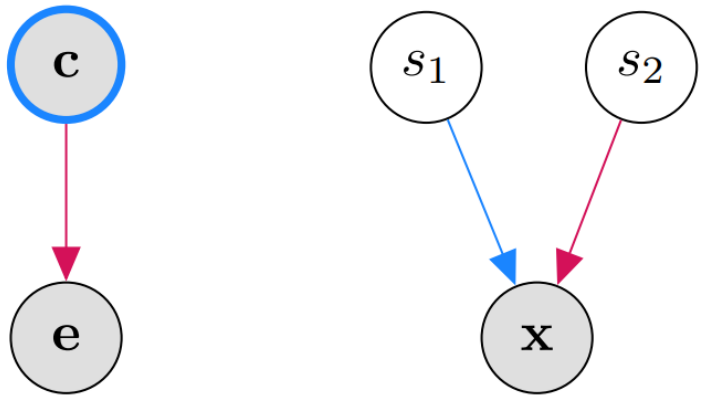
$$\log|J_f(s)| = \sum_{i=1}^{n} log\left(\det\left[\frac{\partial f}{\partial s_i}(s)\right]\right)$$

IMA extends IGCI with and orthogonality condition $\frac{\partial f}{\partial s_i}$ on the columns of a Jacobian:



Again Comparing ICM & IMA

For n=2, two sources:



IMA: Decouples *Cause* and *Mechanism* [Right]

$$\int log\left|J_f\left(s\right)\right| ds = \int \left(\sum_{i=1}^{nlog} \left(\det\left[\frac{\partial\ f}{\partial\ s_i}\left(s\right)\right]\right)\right) p\left(s\right) ds$$

IGCI: Decouples the influence of each *independent component* and *Mechanism* [Left]

$$\int log\left|J_f\left(s\right)\right| p\left(s\right)\ ds\ =\ \int log\left|J_f\left(s\right)\right| ds$$

IMA enforces independence between the contributions of different sources s_i to the mixing function f as captured by f/s_i.

The IMA Contrast and Learning Function

Constructing **Contrast function** and then the **Learning function** for unsupervised decomposition:

The IMA Contrast C_{IMA}

$$C_{IMA}(f,\ p(s)) =$$
$$E_s\left(\sum_{i=1}^{n}\left(log\left|\frac{\partial\ f}{\partial\ s_i}(s)\right|\right) -\ log\left(det\left(J_f(s)\right)\right)\right) =$$
$$E_s\left(log\left(\prod_{i=1}^{n}\left(\left|\frac{\partial f}{\partial\ s_i}(s)\right|\right)\right) -\ log\left(det\left(J_f(s)\right)\right)\right)$$

Regularized maximum-likelihood objective C_{IMA}

Lagrange multiplier for constrained optimzation:

$$L(g;\ x) =\ E_x\left[log\ p_g(x)\right] - \lambda\ C_{IMA}\left(g^{-1},\ p_y\right)$$

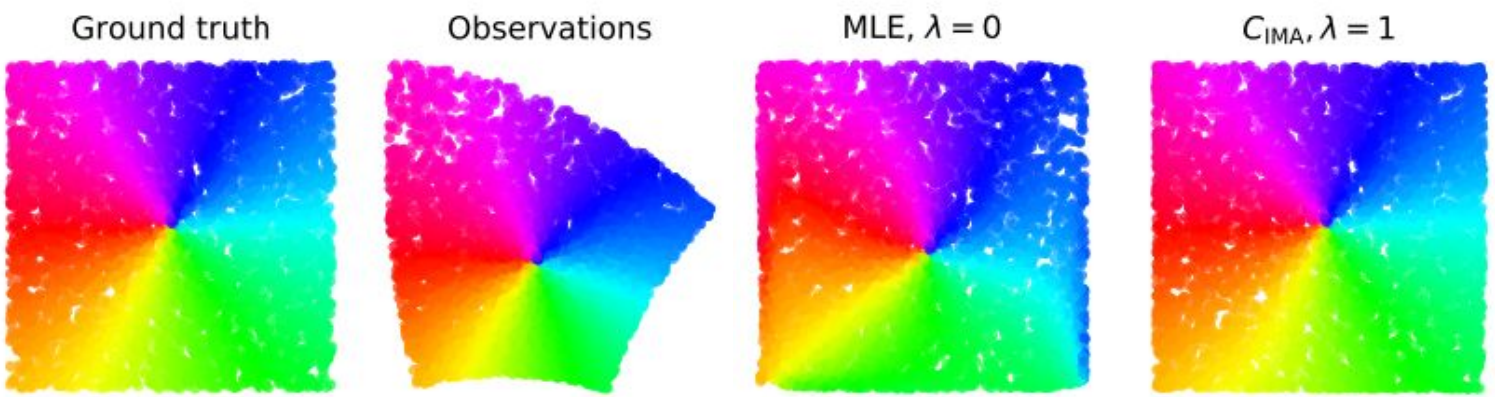
Where g is the learnt unmixing and y = g(x) the reconstructed sources.

Results

Optimizing:

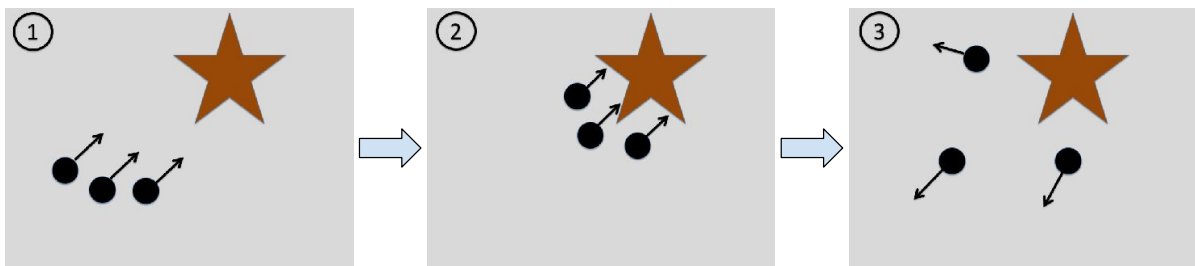
$$L(g;\ x) = \ E_x\left[log\ p_g\left(x\right)\right] - \lambda\ C_{IMA}\left(g^{-1},\ p_y\right)$$

Blind source separation is only achieved when $\lambda > 0$.
For $\lambda = 0$ i.e the maximum likelihood estimation (MLE), the learnt decomposition's are spurious/false.



Miscellaneous Information

- Disentanglement in many way is an Inverse problem - ("ill posed")
- Additional constrains through prior known knowledge might very likely lead to better decomposition's (as seen iVAE), however this might also lead to misleading solutions



8

[8] Inverse problem, Rüland Angkana

Thank you