Disentangling Identifiable Features from Noisy Data with Structured Nonlinear ICA

Hermanni Hälvä[1] **Sylvain Le Corff**[2] **Luc Lehéricy**[3]
Jonathan So[4] **Yongjie Zhu**[1] **Elisabeth Gassiat**[5] **Aapo Hyvärinen**[1]

[1]Department of Computer Science, University of Helsinki
[2]Samovar, Télécom SudParis, Institut Polytechnique de Paris
[3]Laboratoire J. A. Dieudonné, Université Côte d'Azur
[4]Department of Engineering, University of Cambridge
[5]Laboratoire de mathématiques d'Orsay, Université Paris-Saclay

NeurIPS 2021

Overview

- Common assumption in unsupervised representation learning: low-dimensional latent variables generate observed data

Overview

- Common assumption in unsupervised representation learning: low-dimensional latent variables generate observed data
- Knowledge of *true* latent variables would be useful in many tasks: classification, transfer learning, causal inference etc.

Overview

- Common assumption in unsupervised representation learning: low-dimensional latent variables generate observed data
- Knowledge of *true* latent variables would be useful in many tasks: classification, transfer learning, causal inference etc.
- Popular unsupervised approach: learn *disentangled* representation

Overview

- Common assumption in unsupervised representation learning: low-dimensional latent variables generate observed data
- Knowledge of *true* latent variables would be useful in many tasks: classification, transfer learning, causal inference etc.
- Popular unsupervised approach: learn *disentangled* representation
- Problem: models used usually unidentifiable (e.g. β-VAE)

Overview

- Common assumption in unsupervised representation learning: low-dimensional latent variables generate observed data
- Knowledge of *true* latent variables would be useful in many tasks: classification, transfer learning, causal inference etc.
- Popular unsupervised approach: learn *disentangled* representation
- Problem: models used usually unidentifiable (e.g. β-VAE)
- Thus can't recover *true* data generating features

Overview

- Common assumption in unsupervised representation learning: low-dimensional latent variables generate observed data
- Knowledge of *true* latent variables would be useful in many tasks: classification, transfer learning, causal inference etc.
- Popular unsupervised approach: learn *disentangled* representation
- Problem: models used usually unidentifiable (e.g. β-VAE)
- Thus can't recover *true* data generating features
- Our paper: general identifiable framework for principled disentanglement – Structured Nonlinear ICA

Background: identifiability problem

Identifiability

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \implies \boldsymbol{\theta} = \hat{\boldsymbol{\theta}} \;\; \forall(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$$

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$

Background: identifiability problem

Identifiability

$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \implies \boldsymbol{\theta} = \hat{\boldsymbol{\theta}} \ \forall(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$

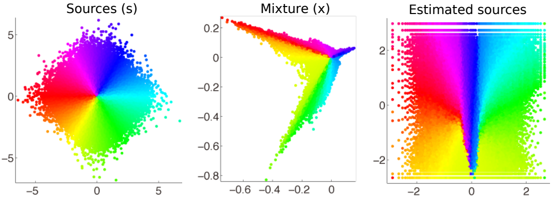
- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- ..are unidentifiable with factorial prior: $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (Hyvärinen and Pajunen, 1999; Khemakhem et al., 2020; Locatello et al., 2019)

Background: identifiability problem

Identifiability

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \implies \boldsymbol{\theta} = \hat{\boldsymbol{\theta}} \ \forall(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$$

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- ..are unidentifiable with factorial prior: $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (Hyvärinen and Pajunen, 1999; Khemakhem et al., 2020; Locatello et al., 2019)
- Thus basic VAEs, GANs, Nonlinear ICA etc. are unidentifiable:



Credit: Hiroshi Morioka

Background: identifiability problem

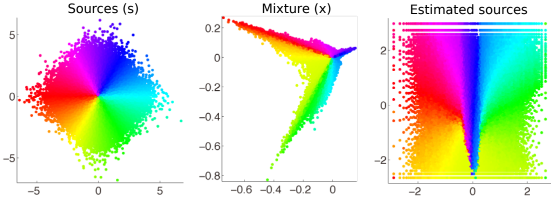
Identifiability

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \implies \boldsymbol{\theta} = \hat{\boldsymbol{\theta}} \ \forall(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$$

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- ..are unidentifiable with factorial prior: $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (Hyvärinen and Pajunen, 1999; Khemakhem et al., 2020; Locatello et al., 2019)
- Thus basic VAEs, GANs, Nonlinear ICA etc. are unidentifiable:

Identifiability problem

$$p_{\mathbf{f}}(\mathbf{x}) = p_{\hat{\mathbf{f}}}(\mathbf{x}) \not\Longrightarrow \mathbf{f} = \hat{\mathbf{f}}$$



Credit: Hiroshi Morioka

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)
- u is *observed* auxiliary variable capturing data structure:

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)
- u is *observed* auxiliary variable capturing data structure:
  - Nonstationarity time-series: $u_t = i, i \in \{1, .., K\}$ indexes non-stationary regions (Hyvärinen and Morioka, 2016)

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)
- u is *observed* auxiliary variable capturing data structure:
  - Nonstationarity time-series: $u_t = i, i \in \{1, .., K\}$ indexes non-stationary regions (Hyvärinen and Morioka, 2016)
  - Autocorrelated time-series: $u_t = \mathbf{x}_{t-1}$ (Hyvärinen and Morioka, 2017)

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)
- u is *observed* auxiliary variable capturing data structure:
  - Nonstationarity time-series: $u_t = i, i \in \{1, .., K\}$ indexes non-stationary regions (Hyvärinen and Morioka, 2016)
  - Autocorrelated time-series: $u_t = \mathbf{x}_{t-1}$ (Hyvärinen and Morioka, 2017)
  - ...or some other observed conditioning variable (class label etc.) Hyvärinen et al. (2019); Khemakhem et al. (2020)

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)
- u is *observed* auxiliary variable capturing data structure:
  - Nonstationarity time-series: $u_t = i, i \in \{1, .., K\}$ indexes non-stationary regions (Hyvärinen and Morioka, 2016)
  - Autocorrelated time-series: $u_t = \mathbf{x}_{t-1}$ (Hyvärinen and Morioka, 2017)
  - ...or some other observed conditioning variable (class label etc.) Hyvärinen et al. (2019); Khemakhem et al. (2020)
- Recently, structure can also be latent (unsupervised):

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)
- u is *observed* auxiliary variable capturing data structure:
  - Nonstationarity time-series: $u_t = i, i \in \{1, .., K\}$ indexes non-stationary regions (Hyvärinen and Morioka, 2016)
  - Autocorrelated time-series: $u_t = \mathbf{x}_{t-1}$ (Hyvärinen and Morioka, 2017)
  - ...or some other observed conditioning variable (class label etc.) Hyvärinen et al. (2019); Khemakhem et al. (2020)
- Recently, structure can also be latent (unsupervised):
  - u_t can be hidden Markov model (Hälvä and Hyvärinen, 2020; Gassiat et al., 2020b)

Background: solving identifiability in Nonlinear ICA

- Deep generative models: $\mathbf{x} = \mathbf{f}(\mathbf{s})$
- Adding structure can recover identifiability:
  - $p(\mathbf{s}) = \prod_{i=1}^{M} p(s_i)$ (unidentifiable)
  - $p(\mathbf{s}|u) = \prod_{i=1}^{M} p(s_i \mid u)$ (possibly identifiable)
- u is *observed* auxiliary variable capturing data structure:
  - Nonstationarity time-series: $u_t = i, i \in \{1, .., K\}$ indexes non-stationary regions (Hyvärinen and Morioka, 2016)
  - Autocorrelated time-series: $u_t = \mathbf{x}_{t-1}$ (Hyvärinen and Morioka, 2017)
  - ...or some other observed conditioning variable (class label etc.) Hyvärinen et al. (2019); Khemakhem et al. (2020)
- Recently, structure can also be latent (unsupervised):
  - u_t can be hidden Markov model (Hälvä and Hyvärinen, 2020; Gassiat et al., 2020b)
- Q: what type of latent structures, in general, allow identifiable disentanglement?

Structured Nonlinear ICA

- General identifiability framework

Structured Nonlinear ICA

- General identifiability framework
- Denote $(\mathbf{x}_t)_{t \in \mathbb{T}} = ((x_t^{(1)}, \ldots, x_t^{(M)}))_{t \in \mathbb{T}}$ where $\mathbb{T}$ is a discrete indexing set of arbitrary dimension.

Structured Nonlinear ICA

- General identifiability framework
- Denote $(\mathbf{x}_t)_{t \in \mathbb{T}} = ((x_t^{(1)}, \ldots, x_t^{(M)}))_{t \in \mathbb{T}}$ where $\mathbb{T}$ is a discrete indexing set of arbitrary dimension.
  - For discrete time-series: $\mathbb{T}$ is a subset of $\mathbb{N}$.

Structured Nonlinear ICA

- General identifiability framework
- Denote $(\mathbf{x}_t)_{t\in\mathbb{T}} = ((x_t^{(1)}, \ldots, x_t^{(M)}))_{t\in\mathbb{T}}$ where $\mathbb{T}$ is a discrete indexing set of arbitrary dimension.
  - For discrete time-series: $\mathbb{T}$ is a subset of $\mathbb{N}$.
  - Crucially, can also be arbitrary indexing variable: e.g. subset of $\mathbb{N}^2$ for spatial data

Structured Nonlinear ICA

- General identifiability framework
- Denote $(\mathbf{x}_t)_{t\in\mathbb{T}} = ((x_t^{(1)}, \ldots, x_t^{(M)}))_{t\in\mathbb{T}}$ where $\mathbb{T}$ is a discrete indexing set of arbitrary dimension.
  - For discrete time-series: $\mathbb{T}$ is a subset of $\mathbb{N}$.
  - Crucially, can also be arbitrary indexing variable: e.g. subset of $\mathbb{N}^2$ for spatial data
- Structured nonlinear ICA (SNICA) assumptions:

Structured Nonlinear ICA

- General identifiability framework
- Denote $(\mathbf{x}_t)_{t\in\mathbb{T}} = ((x_t^{(1)}, \ldots, x_t^{(M)}))_{t\in\mathbb{T}}$ where $\mathbb{T}$ is a discrete indexing set of arbitrary dimension.
  - For discrete time-series: $\mathbb{T}$ is a subset of $\mathbb{N}$.
  - Crucially, can also be arbitrary indexing variable: e.g. subset of $\mathbb{N}^2$ for spatial data
- Structured nonlinear ICA (SNICA) assumptions:
  - Weak stationarity: distributions of $\mathbf{s}_t^{(i)}$ and $\mathbf{s}_{t'}^{(i)}$ are the same for any $t, t' \in \mathbb{T}, \forall i$.

Structured Nonlinear ICA

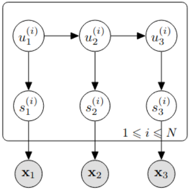
- General identifiability framework
- Denote $(\mathbf{x}_t)_{t \in \mathbb{T}} = ((x_t^{(1)}, \dots, x_t^{(M)}))_{t \in \mathbb{T}}$ where $\mathbb{T}$ is a discrete indexing set of arbitrary dimension.
  - For discrete time-series: $\mathbb{T}$ is a subset of $\mathbb{N}$.
  - Crucially, can also be arbitrary indexing variable: e.g. subset of $\mathbb{N}^2$ for spatial data
- Structured nonlinear ICA (SNICA) assumptions:
  - Weak stationarity: distributions of $\mathbf{s}_t^{(i)}$ and $\mathbf{s}_{t'}^{(i)}$ are the same for any $t, t' \in \mathbb{T}, \forall i$.
  - Unconditional independence of components: $p(\mathbf{s}_{t_1}, \dots, \mathbf{s}_{t_m}) = \prod_{i=1}^{N} p(s_{t_1}^{(i)}, \dots, s_{t_m}^{(i)})$

Structured Nonlinear ICA

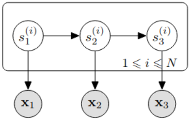
- General identifiability framework
- Denote $(\mathbf{x}_t)_{t\in\mathbb{T}} = ((x_t^{(1)}, \ldots, x_t^{(M)}))_{t\in\mathbb{T}}$ where $\mathbb{T}$ is a discrete indexing set of arbitrary dimension.
  - For discrete time-series: $\mathbb{T}$ is a subset of $\mathbb{N}$.
  - Crucially, can also be arbitrary indexing variable: e.g. subset of $\mathbb{N}^2$ for spatial data
- Structured nonlinear ICA (SNICA) assumptions:
  - Weak stationarity: distributions of $\mathbf{s}_t^{(i)}$ and $\mathbf{s}_{t'}^{(i)}$ are the same for any $t, t' \in \mathbb{T}, \forall i$.
  - Unconditional independence of components: $p(\mathbf{s}_{t_1}, \ldots, \mathbf{s}_{t_m}) = \prod_{i=1}^{N} p(s_{t_1}^{(i)}, \ldots, s_{t_m}^{(i)})$
  - $\mathbf{x}_t = \mathbf{f}(\mathbf{s}_t) + \varepsilon_t$, where ε_t is i.i.d noise with *arbitrary* unknown distribution; $\mathbf{f}$ is injective.

Structured Nonlinear ICA – Examples

- Previous models can be reformulated to fit within our framework



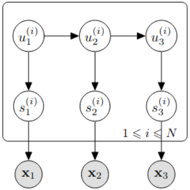
(a) HMM modulated components c.f. (Hälvä and Hyvärinen, 2020)



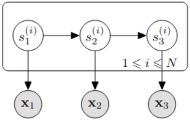
(b) Temporal dependencies c.f. (Hyvärinen and Morioka, 2017)

Structured Nonlinear ICA – Examples

- Previous models can be reformulated to fit within our framework

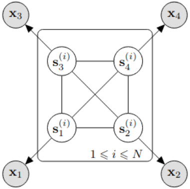


(a) HMM modulated components c.f. (Hälvä and Hyvärinen, 2020)

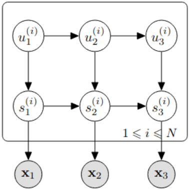


(b) Temporal dependencies c.f. (Hyvärinen and Morioka, 2017)

- As well as flexible new models:



(c) New: Spatial process on a graph (with latent states u_t integrated out)



(d) New: Δ-SNICA , a linear switching dynamics model for components

Identifiability of the SNICA framework (1/3)

- Very general identifiability results for models in SNICA framework

Identifiability of the SNICA framework (1/3)

- Very general identifiability results for models in SNICA framework
- Theorems in two parts:

Identifiability of the SNICA framework (1/3)

- Very general identifiability results for models in SNICA framework
- Theorems in two parts:
  1. Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,

Identifiability of the SNICA framework (1/3)

- Very general identifiability results for models in SNICA framework
- Theorems in two parts:
  1. Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
  2. Identify demixing ($\mathbf{f}^{-1}$) of the nonlinearly mixed data $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:
  - (A1) Tails of $\mathbf{z}_t$ "not much" heavier than Gaussian

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:
  - (A1) Tails of $\mathbf{z}_t$ "not much" heavier than Gaussian
  - (A2) Non-degeneracy assumption

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:
  - (A1) Tails of $\mathbf{z}_t$ "not much" heavier than Gaussian
  - (A2) Non-degeneracy assumption
  - (A3) No direction of $\mathbf{z}_t$ has Gaussian component

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:
  - (A1) Tails of $\mathbf{z}_t$ "not much" heavier than Gaussian
  - (A2) Non-degeneracy assumption
  - (A3) No direction of $\mathbf{z}_t$ has Gaussian component

Theorem

Assume that assumptions (A1), (A2) and (A3) hold for some $(t_1, t_2) \in \mathbb{T}^2$. Then, $\mathbf{z}_t$ is identified, up to translation.

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:
  - (A1) Tails of $\mathbf{z}_t$ "not much" heavier than Gaussian
  - (A2) Non-degeneracy assumption
  - (A3) No direction of $\mathbf{z}_t$ has Gaussian component

Theorem

Assume that assumptions (A1), (A2) and (A3) hold for some $(t_1, t_2) \in \mathbb{T}^2$. Then, $\mathbf{z}_t$ is identified, up to translation.

- Noise ε can have arbitrary and unknown distribution!

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:
  - (A1) Tails of $\mathbf{z}_t$ "not much" heavier than Gaussian
  - (A2) Non-degeneracy assumption
  - (A3) No direction of $\mathbf{z}_t$ has Gaussian component

Theorem

Assume that assumptions (A1), (A2) and (A3) hold for some $(t_1, t_2) \in \mathbb{T}^2$. Then, $\mathbf{z}_t$ is identified, up to translation.

- Noise ε can have arbitrary and unknown distribution!
- Very general result – not limited to our model

Identifiability of the SNICA framework (2/3)

- Identify noise-free distribution of $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$ from noisy data $\mathbf{x}_t = \mathbf{z}_t + \varepsilon_t$,
- Assumptions:
  - (A1) Tails of $\mathbf{z}_t$ "not much" heavier than Gaussian
  - (A2) Non-degeneracy assumption
  - (A3) No direction of $\mathbf{z}_t$ has Gaussian component

Theorem

Assume that assumptions (A1), (A2) and (A3) hold for some $(t_1, t_2) \in \mathbb{T}^2$. Then, $\mathbf{z}_t$ is identified, up to translation.

- Noise ε can have arbitrary and unknown distribution!
- Very general result – not limited to our model
- Extension of Gassiat et al. (2020b,a)

Identifiability of the SNICA framework (3/3)

- Previous theorem gets us to $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$

[1]Can be relaxed under other stricter conditions. See Appendix B.

Identifiability of the SNICA framework (3/3)

- Previous theorem gets us to $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$
- Next, identify demixing $\mathbf{f}^{-1}$

[1] Can be relaxed under other stricter conditions. See Appendix B.

Identifiability of the SNICA framework (3/3)

- Previous theorem gets us to $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$
- Next, identify demixing $\mathbf{f}^{-1}$
- Assumptions

[1] Can be relaxed under other stricter conditions. See Appendix B.

Identifiability of the SNICA framework (3/3)

- Previous theorem gets us to $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$
- Next, identify demixing $\mathbf{f}^{-1}$
- Assumptions
  - (B1) "Sufficient" dependency between "nearby" datapoints for each independent component

[1] Can be relaxed under other stricter conditions. See Appendix B.

Identifiability of the SNICA framework (3/3)

- Previous theorem gets us to $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$
- Next, identify demixing $\mathbf{f}^{-1}$
- Assumptions
  - (B1) "Sufficient" dependency between "nearby" datapoints for each independent component
  - (B2)[1] Distributions of independent components are non-quasi-Gaussian (e.g. no GPs)

[1] Can be relaxed under other stricter conditions. See Appendix B.

Identifiability of the SNICA framework (3/3)

- Previous theorem gets us to $\mathbf{z}_t = \mathbf{f}(\mathbf{s}_t)$
- Next, identify demixing $\mathbf{f}^{-1}$
- Assumptions
  - (B1) "Sufficient" dependency between "nearby" datapoints for each independent component
  - (B2)[1] Distributions of independent components are non-quasi-Gaussian (e.g. no GPs)

Theorem

Assume that assumptions (B1) and (B2) hold, then, $\mathbf{f}^{-1}$ *can be recovered up to permutation and coordinate-wise transformations from the distribution of* $(\mathbf{f}(\mathbf{s}_{t_1}), \ldots, \mathbf{f}(\mathbf{s}_{t_m}))$

[1] Can be relaxed under other stricter conditions. See Appendix B.

New nonlinear ICA model for time-series: Δ-SNICA

- Each independent component follows Switching Linear Dynamical System. For all $i = 1, ..., N$:

$$\mathbf{y}_t^{(i)} = \mathbf{B}_{u_t}^{(i)} \mathbf{y}_{t-1}^{(i)} + \mathbf{b}_{u_t}^{(i)} + \boldsymbol{\varepsilon}_{u_t}^{(i)} , \tag{1}$$

where $u_t := u_t^{(i)}$ is a state of a 1st-order hidden Markov chain, and where the first elements $\mathbf{y}_t^{(i)} = (s_t^{(i)}, \ldots, y_{t,d}^{(i)})^T$, is the ind. comp.

New nonlinear ICA model for time-series: Δ-SNICA

- Each independent component follows Switching Linear Dynamical System. For all $i = 1, ..., N$:

$$\mathbf{y}_t^{(i)} = \mathbf{B}_{u_t}^{(i)} \mathbf{y}_{t-1}^{(i)} + \mathbf{b}_{u_t}^{(i)} + \boldsymbol{\varepsilon}_{u_t}^{(i)}\,, \tag{1}$$

  where $u_t := u_t^{(i)}$ is a state of a 1st-order hidden Markov chain, and where the first elements $\mathbf{y}_t^{(i)} = (s_t^{(i)}, \ldots, y_{t,d}^{(i)})^T$, is the ind. comp.

- $\mathbf{x}_t = \mathbf{f}(s_t^{(1)}, \ldots, s_t^{(N)}) + \boldsymbol{\varepsilon}_t$
  where output noise allows dimensionality reduction

New nonlinear ICA model for time-series: Δ-SNICA

- Each independent component follows Switching Linear Dynamical System. For all $i = 1, ..., N$:

$$\mathbf{y}_t^{(i)} = \mathbf{B}_{u_t}^{(i)} \mathbf{y}_{t-1}^{(i)} + \mathbf{b}_{u_t}^{(i)} + \boldsymbol{\varepsilon}_{u_t}^{(i)}\,, \tag{1}$$

  where $u_t := u_t^{(i)}$ is a state of a 1st-order hidden Markov chain, and where the first elements $\mathbf{y}_t^{(i)} = (s_t^{(i)}, \ldots, y_{t,d}^{(i)})^T$, is the ind. comp.

- $\mathbf{x}_t = \mathbf{f}(s_t^{(1)}, \ldots, s_t^{(N)}) + \boldsymbol{\varepsilon}_t$
  where output noise allows dimensionality reduction
- Accounts for useful properties: autocorrelation, non-stationarity, dimension reduction, and measurement noise.

New nonlinear ICA model for time-series: Δ-SNICA

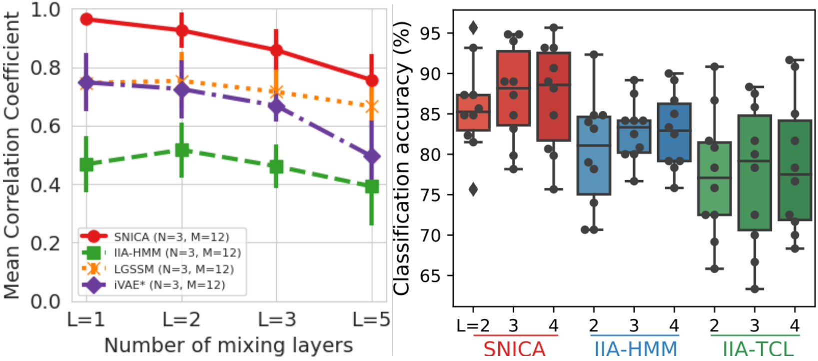
- Each independent component follows Switching Linear Dynamical System. For all $i = 1, ..., N$:

$$\mathbf{y}_t^{(i)} = \mathbf{B}_{u_t}^{(i)} \mathbf{y}_{t-1}^{(i)} + \mathbf{b}_{u_t}^{(i)} + \varepsilon_{u_t}^{(i)} \,, \tag{1}$$

  where $u_t := u_t^{(i)}$ is a state of a 1st-order hidden Markov chain, and where the first elements $\mathbf{y}_t^{(i)} = (s_t^{(i)}, \ldots, y_{t,d}^{(i)})^T$, is the ind. comp.
- $\mathbf{x}_t = \mathbf{f}(s_t^{(1)}, \ldots, s_t^{(N)}) + \varepsilon_t$
  where output noise allows dimensionality reduction
- Accounts for useful properties: autocorrelation, non-stationarity, dimension reduction, and measurement noise.
- Nonlinear ICA for video, audio, financial, brain signal data etc.?

Experiments

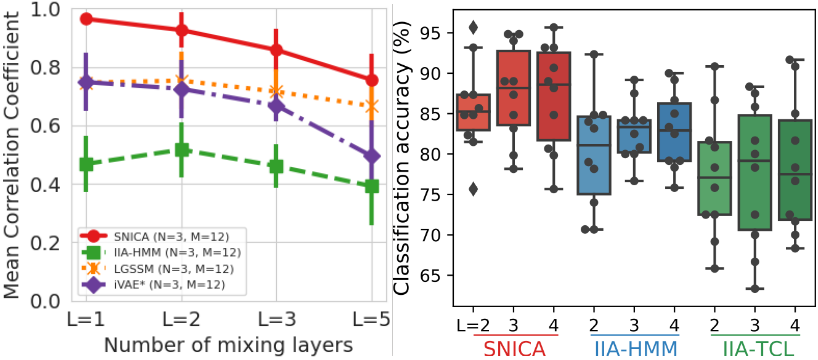
- Estimate Δ-SNICA with variational inference (Structured VAE)



IIA-HMM: independent innovation analysis with hidden markov latent states
IIA-TCL: independent innovation analysis with time-contrastive learning
LGSSM: linear Gaussian state-space model
iVAE*: identifiable VAE with ground-truth HMM state as auxiliary variable

Experiments

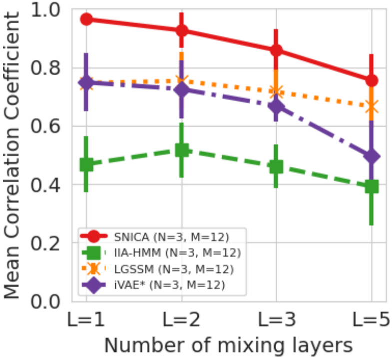
- Estimate Δ-SNICA with variational inference (Structured VAE)
- Simulated data (LHS): Measure identifiability – correlation between estimated and true independent components

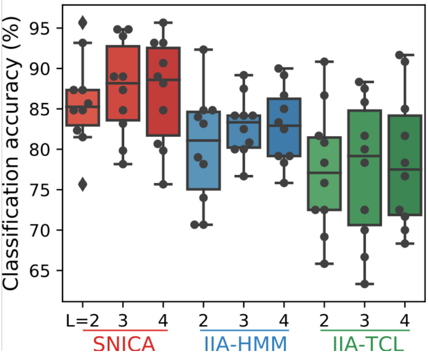


IIA-HMM: independent innovation analysis with hidden markov latent states
IIA-TCL: independent innovation analysis with time-contrastive learning
LGSSM: linear Gaussian state-space model
iVAE*: identifiable VAE with ground-truth HMM state as auxiliary variable

Experiments

- Estimate Δ-SNICA with variational inference (Structured VAE)
- Simulated data (LHS): Measure identifiability – correlation between estimated and true independent components
- MEG data (RHS) – feature extraction and classification of stimulus categories:





IIA-HMM: independent innovation analysis with hidden markov latent states
IIA-TCL: independent innovation analysis with time-contrastive learning
LGSSM: linear Gaussian state-space model
iVAE*: identifiable VAE with ground-truth HMM state as auxiliary variable

Conclusions

- General theoretical identifiability framework

Conclusions

- General theoretical identifiability framework
- Principled disentanglement in structured models by Nonlinear ICA

Conclusions

- General theoretical identifiability framework
- Principled disentanglement in structured models by Nonlinear ICA
- Identifiable deconvolution even when output noise is arbitrary and unknown

Conclusions

- General theoretical identifiability framework
- Principled disentanglement in structured models by Nonlinear ICA
- Identifiable deconvolution even when output noise is arbitrary and unknown
- Δ-SNICA allows for rich temporal dynamics

Conclusions

- General theoretical identifiability framework
- Principled disentanglement in structured models by Nonlinear ICA
- Identifiable deconvolution even when output noise is arbitrary and unknown
- Δ-SNICA allows for rich temporal dynamics
- Multiple new models can be developed e.g. spatial/image data

Conclusions

- General theoretical identifiability framework
- Principled disentanglement in structured models by Nonlinear ICA
- Identifiable deconvolution even when output noise is arbitrary and unknown
- Δ-SNICA allows for rich temporal dynamics
- Multiple new models can be developed e.g. spatial/image data
- Future theoretical work needed for: heavy tails, non-additive output noise, noise that's not independent of the signal.

References

Gassiat, E., Le Corff, S., and Lehéricy, L. (2020a). Deconvolution with unknown noise distribution is possible for multivariate signals. *arXiv:2006.14226*.

Gassiat, E., Le Corff, S., and Lehéricy, L. (2020b). Identifiability and consistent estimation of nonparametric translation hidden markov models with general state space. *Journal of Machine Learning Research*, 21(115):1–40.

Hälvä, H. and Hyvärinen, A. (2020). Hidden Markov nonlinear ICA: Unsupervised learning from nonstationary time series. In *Proc. 36th Conf. on Uncertainty in Artificial Intelligence (UAI2020)*, Toronto, Canada (virtual).

Hyvärinen, A. and Morioka, H. (2016). Unsupervised feature extraction by time-contrastive learning and nonlinear ICA. In *Advances in Neural Information Processing Systems (NIPS2016)*, Barcelona, Spain.

Hyvärinen, A. and Morioka, H. (2017). Nonlinear ICA of temporally dependent stationary sources. In *Proc. Artificial Intelligence and Statistics (AISTATS2017)*, Fort Lauderdale, Florida.