Hidden Markov Nonlinear ICA:

Unsupervised Learning of Independent Components from Nonstationary Time Series

Hermanni Hälvä[1] **Aapo Hyvärinen**[1, 2]

[1]University of Helsinki

[2]Université Paris-Saclay, Inria

UAI 2020, August 3 - 6

Background - ICA

- Learn interesting and useful *latent* representations in data

Background - ICA

- Learn interesting and useful *latent* representations in data
- e.g. independent latent components – 'principled disentanglement'

Background - ICA

- Learn interesting and useful *latent* representations in data
- e.g. independent latent components – 'principled disentanglement'
- In the past – linear ICA:

$$\mathbf{x} = \mathbf{A}\mathbf{s}\,;\, p(\mathbf{s}) = \prod_{i}^{N} p_i(s_i) \tag{1}$$

Background - ICA

- Learn interesting and useful *latent* representations in data
- e.g. independent latent components – 'principled disentanglement'
- In the past – linear ICA:

$$\mathbf{x} = \mathbf{A}\mathbf{s}\ ; p(\mathbf{s}) = \prod_i^N p_i(s_i) \tag{1}$$

- What we really want – *Nonlinear* ICA:

$$\mathbf{x} = \mathbf{f}(\mathbf{s})\ ; p(\mathbf{s}) = \prod_i^N p_i(s_i) \tag{2}$$

Background - ICA

- Learn interesting and useful *latent* representations in data
- e.g. independent latent components – 'principled disentanglement'
- In the past – linear ICA:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \; ; p(\mathbf{s}) = \prod_{i}^{N} p_i(s_i) \tag{1}$$

- What we really want – *Nonlinear* ICA:

$$\mathbf{x} = \mathbf{f}(\mathbf{s}) \; ; p(\mathbf{s}) = \prod_{i}^{N} p_i(s_i) \tag{2}$$

- Problem: nonlinear ICA is non-identifiable! (unsolved problem for decades...)

Background - Identifiability

Identifiability

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

- Without identifiability true representations can't be recovered

Background - Identifiability

Identifiability

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

- Without identifiability true representations can't be recovered

Identifiability problem – nonlinear ICA

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \not\Rightarrow \mathbf{s} = \hat{\mathbf{s}} \;\; \text{where } \mathbf{s} = \mathbf{f}_{\boldsymbol{\theta}}^{-1}(\mathbf{x})$$

Background - Identifiability

Identifiability

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

- Without identifiability true representations can't be recovered

Identifiability problem – nonlinear ICA

$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \not\Rightarrow \mathbf{s} = \hat{\mathbf{s}}$ where $\mathbf{s} = \mathbf{f}_{\boldsymbol{\theta}}^{-1}(\mathbf{x})$

- *Unconditional* factorial prior on latent variable $\implies$ unidentifiable latent variable model (c.f. VAEs, GANs)

Background - Identifiability

Identifiability

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

- Without identifiability true representations can't be recovered

Identifiability problem – nonlinear ICA

$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \not\Rightarrow \mathbf{s} = \hat{\mathbf{s}}$ where $\mathbf{s} = \mathbf{f}_{\boldsymbol{\theta}}^{-1}(\mathbf{x})$

- *Unconditional* factorial prior on latent variable $\Longrightarrow$ unidentifiable latent variable model (c.f. VAEs, GANs)
- Identifiability requires *conditional* (factorial) prior:

$$p(\mathbf{s}) = \prod p(s_i) \rightarrow p(\mathbf{s}|c) = \prod p(s_i|c) \tag{3}$$

Background - Identifiability

Identifiability

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

- Without identifiability true representations can't be recovered

Identifiability problem – nonlinear ICA

$p_{\boldsymbol{\theta}}(\mathbf{x}) = p_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) \not\Rightarrow \mathbf{s} = \hat{\mathbf{s}}$ where $\mathbf{s} = \mathbf{f}_{\boldsymbol{\theta}}^{-1}(\mathbf{x})$

- *Unconditional* factorial prior on latent variable $\implies$ unidentifiable latent variable model (c.f. VAEs, GANs)
- Identifiability requires *conditional* (factorial) prior:

$$p(\mathbf{s}) = \prod p(s_i) \rightarrow p(\mathbf{s}|c) = \prod p(s_i|c) \tag{3}$$

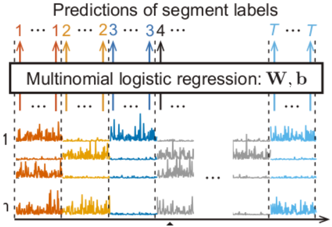
- see e.g. Khemakhem et al. (2020)

Identifiable Nonlinear ICA

- First identifiable nonlinear ICA was Time Contrastive Learning (Hyvärinen and Morioka, 2016)

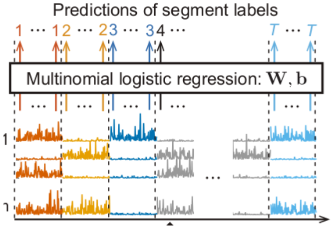
Identifiable Nonlinear ICA

- First identifiable nonlinear ICA was Time Contrastive Learning (Hyvärinen and Morioka, 2016)
- TCL has $p(\mathbf{s}|c) = \prod_i p(s_i|c)$, where c identifies different non-stationary segments:



Identifiable Nonlinear ICA

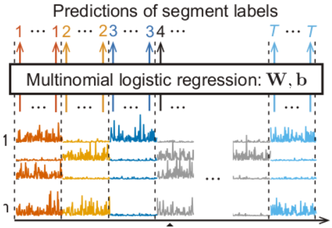
- First identifiable nonlinear ICA was Time Contrastive Learning (Hyvärinen and Morioka, 2016)
- TCL has $p(\mathbf{s}|c) = \prod_i p(s_i|c)$, where c identifies different non-stationary segments:



- **Problems**:

Identifiable Nonlinear ICA

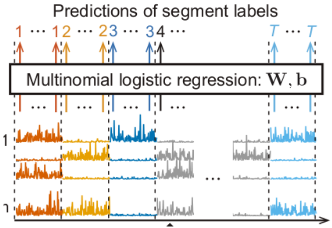
- First identifiable nonlinear ICA was Time Contrastive Learning (Hyvärinen and Morioka, 2016)
- TCL has $p(\mathbf{s}|c) = \prod_i p(s_i|c)$, where c identifies different non-stationary segments:



- **Problems**:
  - Assumes non-stationary segments indices are *observed* – not truly unsupervised

Identifiable Nonlinear ICA

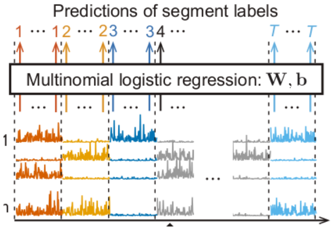
- First identifiable nonlinear ICA was Time Contrastive Learning (Hyvärinen and Morioka, 2016)
- TCL has $p(\mathbf{s}|c) = \prod_i p(s_i|c)$, where c identifies different non-stationary segments:



- **Problems**:
  - Assumes non-stationary segments indices are *observed* – not truly unsupervised
  - For large data, segment indices impossible to find manually

Identifiable Nonlinear ICA

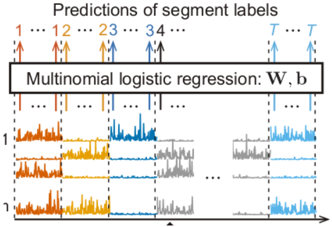
- First identifiable nonlinear ICA was Time Contrastive Learning (Hyvärinen and Morioka, 2016)
- TCL has $p(\mathbf{s}|c) = \prod_i p(s_i|c)$, where c identifies different non-stationary segments:



- **Problems**:
  - Assumes non-stationary segments indices are *observed* – not truly unsupervised
  - For large data, segment indices impossible to find manually
  - In practice segment at equal intervals

Identifiable Nonlinear ICA

- First identifiable nonlinear ICA was Time Contrastive Learning (Hyvärinen and Morioka, 2016)
- TCL has $p(\mathbf{s}|c) = \prod_i p(s_i|c)$, where c identifies different non-stationary segments:



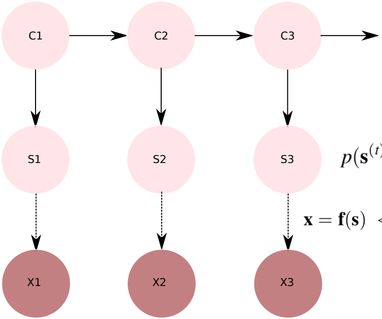
- **Problems**:
  - Assumes non-stationary segments indices are *observed* – not truly unsupervised
  - For large data, segment indices impossible to find manually
  - In practice segment at equal intervals
  - Inaccurate, unprincipled, ignores temporal latent dynamics

Solution

- Learn states and dynamics in a hidden Markov model (HMM) framework!

Solution

- Learn states and dynamics in a hidden Markov model (HMM) framework!

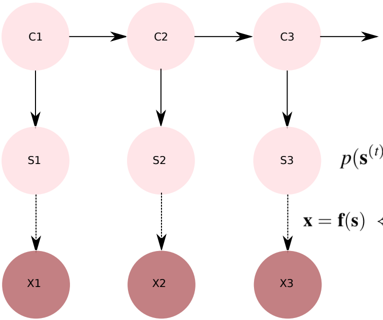


$$p(\mathbf{s}^{(t)}|c^{(t)};\boldsymbol{\lambda}_{c^{(t)}}) = \prod_{i=1}^{N} \frac{h(s_i^{(t)})}{Z(\boldsymbol{\lambda}_{i,c^{(t)}})} \exp\{\langle \boldsymbol{\lambda}_{i,c^{(t)}}, \mathbf{T}_i(s_i)\rangle\}$$

$$\mathbf{x} = \mathbf{f}(\mathbf{s}) \iff \mathbf{g}(\mathbf{x}) = \mathbf{s}$$

Solution

- Learn states and dynamics in a hidden Markov model (HMM) framework!



$$p(\mathbf{s}^{(t)}|c^{(t)};\boldsymbol{\lambda}_{c^{(t)}}) = \prod_{i=1}^{N} \frac{h(s_i^{(t)})}{Z(\boldsymbol{\lambda}_{i,c^{(t)}})} \exp\{\langle \boldsymbol{\lambda}_{i,c^{(t)}}, \mathbf{T}_i(s_i)\rangle\}$$

$$\mathbf{x} = \mathbf{f}(\mathbf{s}) \iff \mathbf{g}(\mathbf{x}) = \mathbf{s}$$

- Learning by EM (Baum-Welch) – including posteriors of conditioning variable

Results – identifiability theory

- Central question: can **latent** conditioning variable identify nonlinear ICA?

Results – identifiability theory

- Central question: can **latent** conditioning variable identify nonlinear ICA?
- Gassiat et al. (2016) – state conditional output distributions of any HMM are identified (some assumptions)

Results – identifiability theory

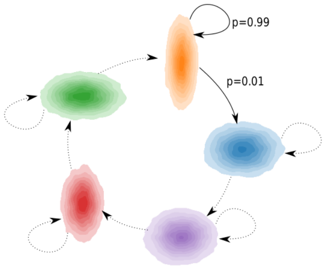
- Central question: can **latent** conditioning variable identify nonlinear ICA?
- Gassiat et al. (2016) – state conditional output distributions of any HMM are identified (some assumptions)
- Our paper – latent state temporal structure identifies ICs via HMM identifiability (proof in paper):

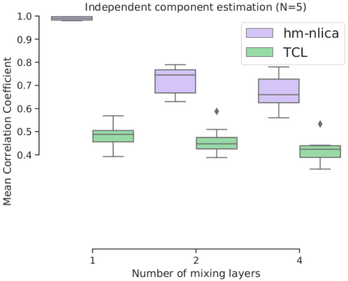
Results – identifiability theory

- Central question: can **latent** conditioning variable identify nonlinear ICA?
- Gassiat et al. (2016) – state conditional output distributions of any HMM are identified (some assumptions)
- Our paper – latent state temporal structure identifies ICs via HMM identifiability (proof in paper):
  - Strong identifiability (under some assumptions):

$$s_i = w_{ij}\hat{g}_j(\mathbf{x}) + b_{ij}$$

Results – simulations





Conclusion

- First unsupervised identifiable nonlinear ICA – *latent* conditioning variable is learned by HMM

Conclusion

- First unsupervised identifiable nonlinear ICA – *latent* conditioning variable is learned by HMM
- Strong identifiability results

Conclusion

- First unsupervised identifiable nonlinear ICA – *latent* conditioning variable is learned by HMM
- Strong identifiability results
- Many real world data, e.g. brain imaging and video, are non-stationary and will be experimented upon

Conclusion

- First unsupervised identifiable nonlinear ICA – *latent* conditioning variable is learned by HMM
- Strong identifiability results
- Many real world data, e.g. brain imaging and video, are non-stationary and will be experimented upon
- More complex conditional independence structures, dimension reduction, are some theoretical extension in progress

References

Gassiat, E., Cleynen, A., and Robin, S. (2016). Inference in finite state space non parametric Hidden Markov Models and applications. *Statistics and Computing*, 26(1-2):61–71.

Hyvärinen, A. and Morioka, H. (2016). Unsupervised Feature Extraction by Time-Contrastive Learning and Nonlinear ICA. *arXiv:1605.06336 [cs, stat]*. arXiv: 1605.06336.

Khemakhem, I., Kingma, D. P., Monti, R. P., and Hyvärinen, A. (2020). Variational Autoencoders and Nonlinear ICA: A Unifying Framework. *arXiv:1907.04809 [cs, stat]*. arXiv: 1907.04809.