NEURAL INFORMATION PROCESSING SYSTEMS

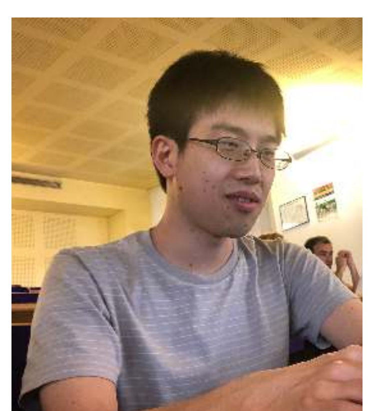
Coupling-based Invertible Neural Networks Are Universal Diffeomorphism Approximators

Takeshi Teshima[1 2], Isao Ishikawa[3 2], Koichi Tojo[2], Kenta Oono[1], Masahiro Ikeda[2], Masashi Sugiyama[2 1]

[1]The University of Tokyo, Japan [2]RIKEN, Japan [3] Ehime University, Japan

NeurIPS2020






Target audience & structure

Part 1

For everyone.

Part 2

For INN users.

Part 3

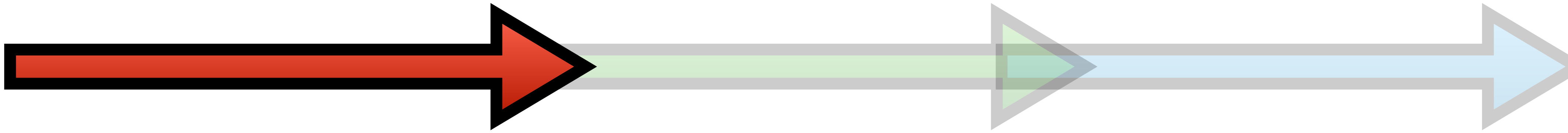
For those interested in details.

Disclaimer

Many descriptions are informal. Please see paper for precise info.

Sec. ? indicates the corresponding section of our paper.

Part 1



Part 1
For everyone.

Part 2
For INN users.

Part 3
For those interested in details.

- What we did and why we did it.

Take-home message

What did we do?

Theoretically investigated:
Are our INNs expressive enough?

INNs = Invertible neural networks

Why important?

Models without a representation power guarantee are hard to rely on.

What is the result?

"Coupling-based INNs (CF-INNs)" are "universal function approximators" despite their restricted architectures.

Message

"CF-INNs" can be relied on in modelling invertible functions and probability distributions.

What is "coupling-based INN?"

Definition (informal)

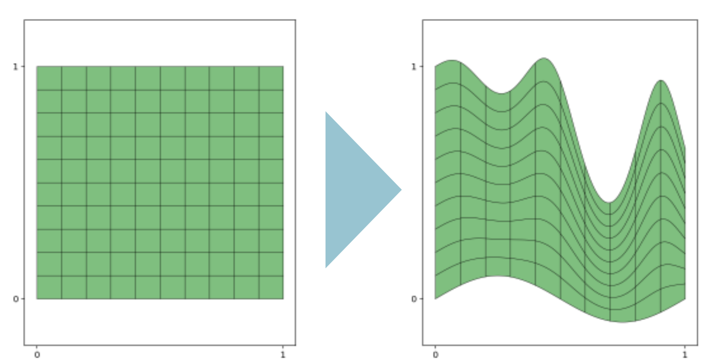
Sec. 2.1

Invertible neural network (INN) is a finite composition of invertible **affine transforms** and invertible **flow layers**.

Coupling flows [DKB14, PNRML19, KPB19]

Sec. 2.1

$$\begin{bmatrix} x_{\le k} \\ x_{>k} \end{bmatrix} \mapsto \begin{bmatrix} x_{\le k} \\ \mathcal{T}(x_{>k},\ \theta(x_{\le k})) \end{bmatrix}$$



Idea: Keep some dimensions unchanged.

CF-INN = Coupling-flow based INN.

Research question

Research question

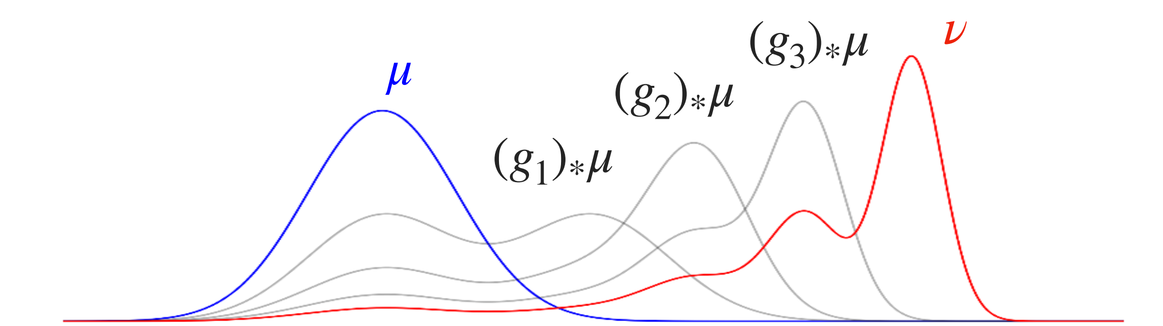
Can CF-INNs have sufficient representation power?

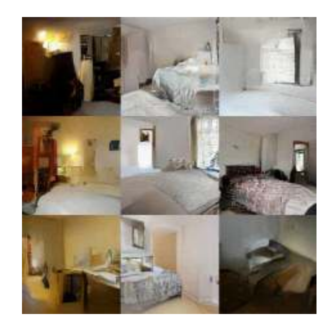
(Restricted function form → restricted representation power?)

Usages of CF-INNs

Sec. 1

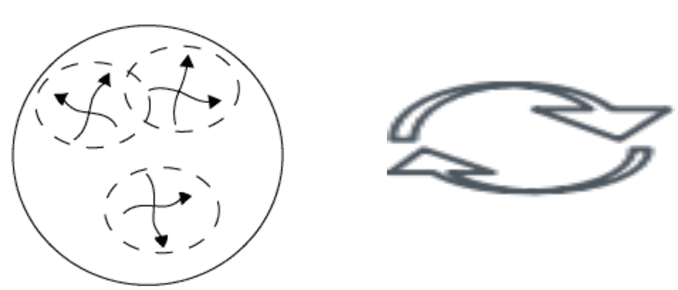
- Approximate distributions (normalizing flows).





[KD18]

- Approximate invertible maps (feature extraction & manipulation).



[DSB17]

Short answer is yes

Research question

Can CF-INNs have sufficient representation power?

(Restricted function form $\to$ restricted representation power?)

Answer

Yes.

Part 2

Part 1

For everyone.

Part 2

For INN users.

Part 3

For those interested in details.

- Technically, what our results are and what they mean.

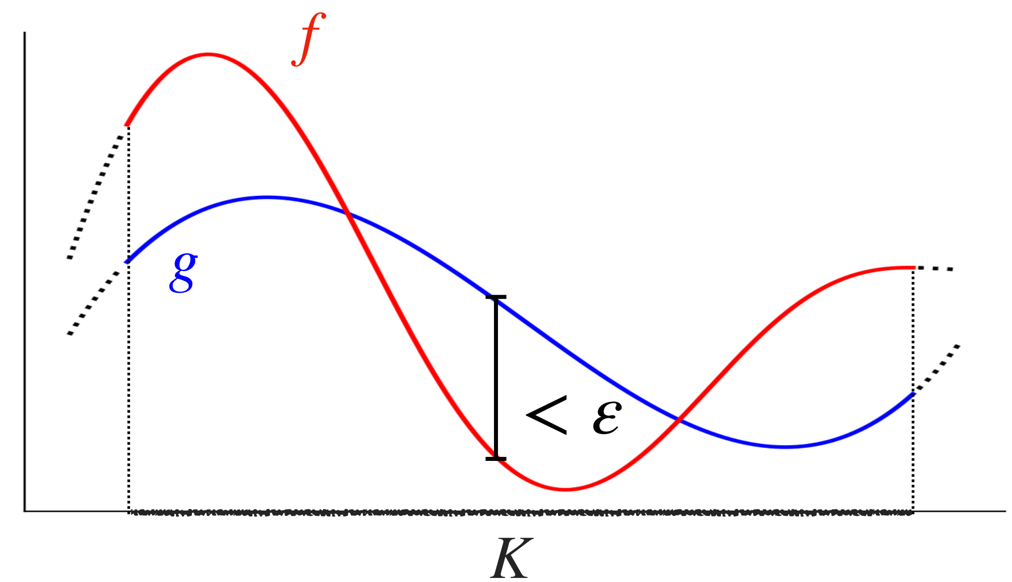
What is "representation power"?

"Representation power" = Universal approximation property

Definition (informal) [C89,HSW89]

Sec. 2.2

$\sup$**- (**L^p**-) universal approximator**: the model can approximate any target function w.r.t. $\sup$- (L^p-) norm on a compact set.



Approximation Target $\mathscr{D}^2$

Sec. 3.1

Fairly **large set** of smooth invertible maps.

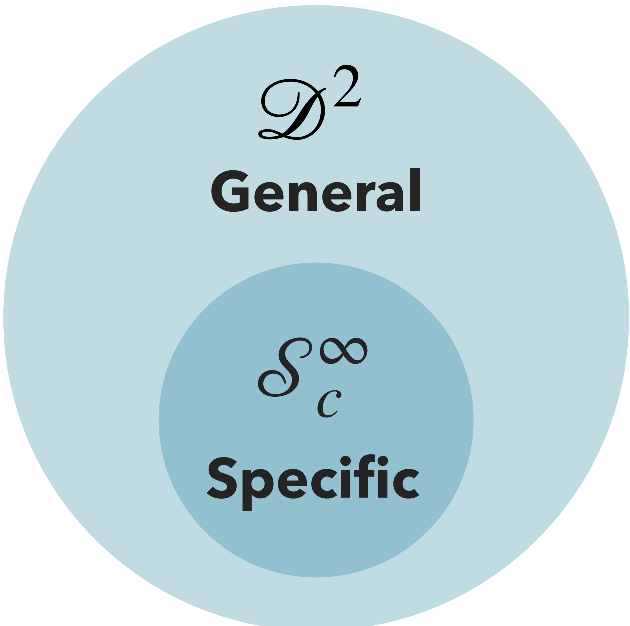
$\mathscr{D}^2$
General
$\{$Diffeo on $\mathbb{R}^d\}$
...and more

Result 1: CF-INNs are universal

Theorem (informal) — **Sec. 3.1, Theorem 1**

$\sup$- (L^p-) **universal approximator for** $\mathscr{S}_c^\infty$

$\Longrightarrow$ $\sup$- (L^p-) **universal approximator for** $\mathscr{D}^2$.



Application

We used the result to demonstrate that

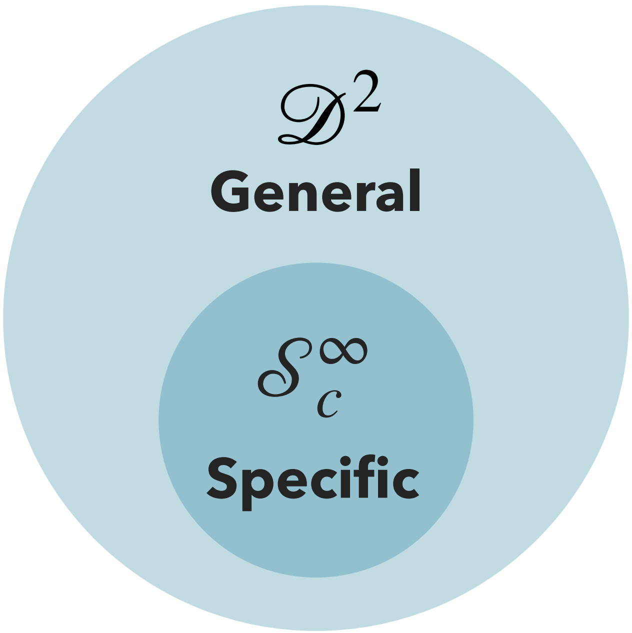
- Sum-of-squares polynomial flow (SoS-flow) [JSY19]
- Deep sigmoidal flow (DSF; aka. NAF) [HKLC18]

yield $\sup$-universal CF-INNs **for** $\mathscr{D}^2$ (stronger than in [JSY19, HKLC18]).

(Advanced) How Result 1 can be used

You show		You get
sup-univ. for $\mathscr{S}_c^\infty$	$\Longrightarrow$	sup-univ. for $\mathscr{D}^2$
$\Downarrow$		$\Downarrow$
L^p-univ. for $\mathscr{S}_c^\infty$	$\Longrightarrow$	L^p-univ. for $\mathscr{D}^2$

Sec. 3.1



$\mathscr{S}_c^\infty := \{\tau :$ compactly supported $\tau(\mathbf{x}, y) = (\mathbf{x}, u(\mathbf{x}, y))\}$

Affine-coupling flows (ACFs)

Definition (informal) [DKB14,DSB17,KD18]

Sec. 2.1

(Single-coordinate) **Affine coupling flows** (ACFs) is a special CF architecture:

$$\Psi_{s,t}(\mathbf{x}, y) := \left(\mathbf{x}, e^{s(\mathbf{x})}y + t(\mathbf{x})\right)$$

Why are ACFs interesting?


[DSB17]

- Popular in applications
  - Generative modeling [DSB17,KD18,OLB+18,KLSKY19,ZMWN19]
  - Probabilistic inference [BM19,WSB19,LW17,AKRK19]
  - Semi-supervised learning [IKFW20]
  - Transfer learning [TSS20], etc.
- Simplest architecture

  → Theoretical guarantee for ACFs apply to more complex CFs.

Result 2: ACF-INN is L^p-/dist. universal

Theorem (informal)

Sec. 3.2, Theorems 2, 3

ACF-INN is an L^p-universal approximator for $\mathscr{D}^2$.

(As a result,) ACF-INN is a distributional universal approximator.

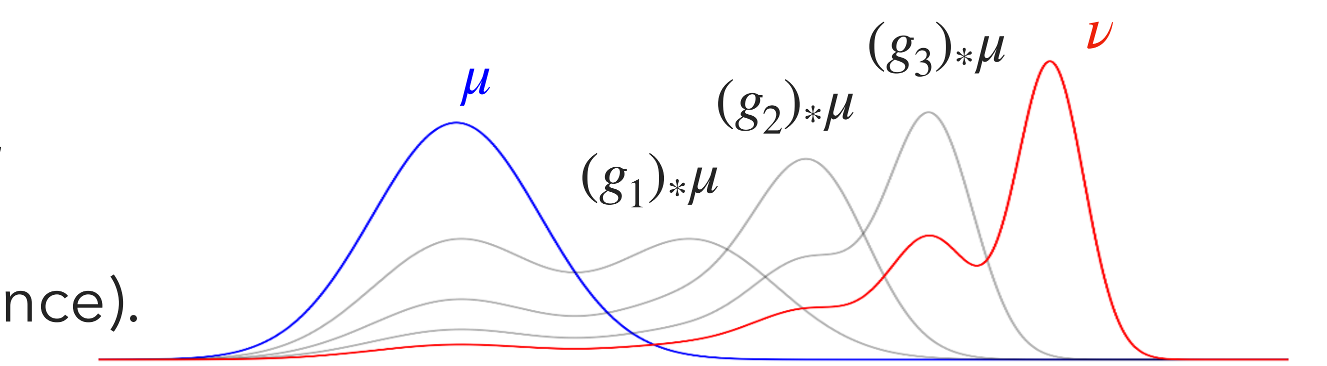
Definition (informal)

Sec. 2.2

A model is a **distributional universal approximator** if it can transform one distribution arbitrarily close to any distribution.

$$(g_n)_*\mu \xrightarrow[n\to\infty]{} \nu$$

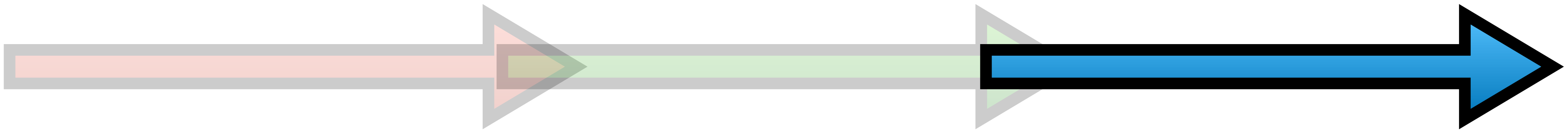
(**weak** convergence).



Implication

- Useful criterion: "if my CF architecture contains ACFs (as special cases), then they are also (L^p-/dist.) universal."
- Affirmative answer to an unsolved conjecture.

Part 3



Part 1
For everyone.

Part 2
For INN users.

Part 3
For those interested in details.

- How Result 1 was obtained.

Proof outline of Result 1

$f \in \mathscr{D}^2$: target, $K \subset U_f$: compact

$f|_K$

$\|$

$\exists W \circ h$ (Aff & compactly supported C^2-diffeomorphism)

$\|$ « **structure theorem of diffeomorphism group**

$\exists h_1 \circ h_2 \circ \cdots$ (**flow endpoints**)

$\|$

$\exists g_1 \circ g_2 \circ \cdots$ (nearly Ids)

$\|$

$\tau_1 \circ \sigma_1 \circ \cdots$ (**permutations & $\mathscr{S}_c^\infty$**)

Decompose $f|_K$ into simpler mappings

Sec. 4

Take-home message

What did we do? Theoretically investigated:
Are our INNs expressive enough?

Why important? Models without a representation power guarantee are hard to rely on.

What is the result? "Coupling-based INNs (CF-INNs)" are "universal function approximators" despite their restricted architectures.

Message

"CF-INNs" can be relied on in modelling invertible functions and probability distributions.

References

[C89] Cybenko, G. (1989).
Approximation by superpositions of a sigmoidal function.
Mathematics of Control, Signals, and Systems, 2, 303–314.

[HSW89] Hornik, K., Stinchcombe, M., & White, H. (1989).
Multilayer feedforward networks are universal approximators.
Neural Networks, 2(5), 359–366.

[JSY19] Jaini, P., Selby, K. A., & Yu, Y. (2019).
Sum-of-squares polynomial flow.
Proceedings of the 36th International Conference on Machine Learning, 97, 3009–3018.

[HKLC18] Huang, C.-W., Krueger, D., Lacoste, A., & Courville, A. (2018).
Neural autoregressive flows.
Proceedings of the 35th International Conference on Machine Learning, 80, 2078–2087.

[KD18] Kingma, D. P., & Dhariwal, P. (2018).
Glow: Generative flow with invertible 1x1 convolutions.
In Advances in Neural Information Processing Systems 31 (pp. 10215–10224).

[PNRML19] Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2019).
Normalizing flows for probabilistic modeling and inference.
ArXiv:1912.02762 [Cs, Stat].

[KPB19] Kobyzev, I., Prince, S., & Brubaker, M. A. (2019).
Normalizing flows: An introduction and review of current methods.
ArXiv:1908.09257 [Cs, Stat].

References

[DKB14] Dinh, L., Krueger, D., & Bengio, Y. (2014).
NICE: Non-linear independent components estimation.
ArXiv:1410.8516 [Cs.LG].

[DSB17] Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2017).
Density estimation using Real NVP.
Fifth International Conference on Learning Representations (ICLR)

[AKRK19] Ardizzone, L., Kruse, J., Rother, C., & Köthe, U. (2019).
Analyzing inverse problems with invertible neural networks.
7th International Conference on Learning Representations.

[BM19] Bauer, M., & Mnih, A. (2019).
Resampled priors for variational autoencoders.
In Proceedings of machine learning research, 89, 66–75.

[LW17] Louizos, C., & Welling, M. (2017).
Multiplicative normalizing flows for variational Bayesian neural networks.
In Proceedings of the 34th International Conference on Machine Learning,70, 2218–2227.

[NMT+19] Nalisnick, E. T., Matsukawa, A., Teh, Y. W., Görür, D., & Lakshminarayanan, B. (2019).
Hybrid models with deep and invertible features.
In Proceedings of the 36th International Conference on Machine Learning, 97, 4723–4732.

[IKFW20] Izmailov, P., Kirichenko, P., Finzi, M., & Wilson, A. G. (2020).
Semi-supervised learning with normalizing flows.
Proceedings of the 37th International Conference on Machine Learning.

References

[KD18] Kingma, D. P., & Dhariwal, P. (2018).
Glow: Generative flow with invertible 1x1 convolutions.
In Advances in Neural Information Processing Systems 31, 10215–10224.

[OLB+18] Oord, A., Li, Y., Babuschkin, I., Simonyan, K., Vinyals, O., Kavukcuoglu, K., Driessche, G., Lockhart, E., Cobo, L., Stimberg, F., Casagrande, N., Grewe, D., Noury, S., Dieleman, S., Elsen, E., Kalchbrenner, N., Zen, H., Graves, A., King, H., … Hassabis, D. (2018).
Parallel WaveNet: Fast high-fidelity speech synthesis.
Proceedings of the 35th International Conference on Machine Learning, 80, 3918–3926.

[TSS20] Teshima, T., Sato, I., & Sugiyama, M. (2020).
Few-shot domain adaptation by causal mechanism transfer.
Proceedings of the 37th International Conference on Machine Learning.

[KLSKY19] Kim, S., Lee, S.-G., Song, J., Kim, J., & Yoon, S. (2019).
FloWaveNet: A generative flow for raw audio.
In Proceedings of the 36th International Conference on Machine Learning, 97, 3370–3378.

[ZMWN19] Zhou, C., Ma, X., Wang, D., & Neubig, G. (2019).
Density matching for bilingual word embedding.
Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 1588–1598.

[WSB19] Ward, P. N., Smofsky, A., & Bose, A. J. (2019).
Improving exploration in soft-actor-critic with normalizing flows policies.
ArXiv:1906.02771 [Cs, Stat].