Relaxing Bijectivitiy Constraints with Continuously Indexed Normalising Flows

ICML 2020

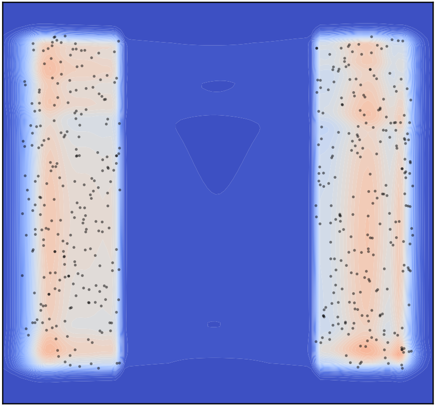
Rob Cornish, Anthony Caterini, George Deligiannidis, Arnaud Doucet

University of Oxford

July 12-18, 2020

Motivation

The following densities were learned using a Gaussian prior with a 10-layer Residual Flow [Chen et al., 2019] (.5M parameters) trained to convergence.



S_X«}0 oI K9}c0} }0X_nj _jL_H9§0 en°0} L0j _§´u K9§9 \n°j _j @e9Hcu

Why Does This Occur?

Normalising Flows (NFs) define the following process:

$$\mu \quad p_{\mu}; \qquad \pm := T(\mu);$$

where T is a diffeomorphism.

Why Does This Occur?

Normalising Flows (NFs) define the following process:

$$\mu \qquad p_{\mu}; \qquad \pm := T(\mu);$$

where T is a diffeomorphism.

Hence the support of $\pm$ will share the same topological properties as the support of μ, i.e.

- Number of connected components
- Number of “holes”
- How they are “knotted”
- etc.

Problem

This suggests a problem when the support of the prior p_μ is simple (e.g. a Gaussian): we usually can't then reproduce the target exactly.

Problem

This suggests a problem when the support of the prior p_μ is simple (e.g. a Gaussian): we usually can't then reproduce the target exactly.

Moreover, to approximate the target closely, our flow must approach non-invertibility.

Our Proposal: Continuously Indexed Flows

Continuously indexed flows (CIFs) instead use the process

$$\mu \quad p_{\mu}; \qquad {}^{a}\, j\, \mu \quad p_{{}^{a} j \mu}(\ j\ \mu); \qquad \pm := S(\mu;\ {}^{a});$$

where ª is a continuous index variable, and each $S(\ ;«)$ is a normalising flow.

Our Proposal: Continuously Indexed Flows

Continuously indexed flows (CIFs) instead use the process

$$\mu \quad p_{\mu}; \qquad {}^{a}\, j\, \mu \quad p_{{}^{a} j \mu}(\ j\ \mu); \qquad \pm := S(\mu;\ {}^{a});$$

where ^{a} is a continuous index variable, and each $S(\ ;\ «)$ is a normalising flow.

Any existing normalising flow can be used to construct S.

Pm` S`QTQb H, *QMiBMmQmbHv AM

*QMiBMmQmbHv BM 6/2t2/ ~Qr b mb2 i?2 T`Q+2bb

w S_{w};   l j w $S_{l j w}$(j w);   s := 6(w; l);

r?2`2 Bb +QMiBMmQmb BM 6/2t p( ;m#B2  MQ`K HBbBM

Mv 2tBbiBM; MQ`K HBbBM; ~Qr + 6M #2 mb2/ iQ +QM

+QMiBMmQmb BM/2t K2 bMbiMQ/2 HQbMB;2 Q`7+i #H2-
i` BM2/ pB  M im` H 1G"P Q#D2+iBp2 BMbi2 /X

"2M2}ib

AMimBiBp2Hv- *A6b + M ó+H2 M mTô K bb i? i rQmH bBM;H2 #BD2+iBQMX

6B;m`2 2R k@H v2` _2bB/m H 6HQr UiQTV M/ *QMiBMmQmbI U#QiiQKVX "Qi? mb2 X8J T ` K2i2`bX

:QBM; .22T2`

q? i? TT2Mb r?2M r2 KQ/2H +QKTHB+ i2/ i `;2i mb

h?2Q`2K 7, i?2 Tr`?B0Q`MQM@?QK2QKQ`T?B+ bm TT2QMi b2[m2M+2 Q7 (7M) +Qs bBM /Bbi`B#miBQM QMHv B7

$$\max \operatorname{Lip} 7_M ; \operatorname{Lip} 7_M^R ! 1 :$$

q? i? TT2Mb r?2M r2 KQ/2H +QKTHB+ i2/ i `;2i mb

h?2Q`2K 7, i?2 w`?B0Q`MQM@?QK2QKQ`T?B+ bm T?T2QM i

b2[m2M+2 Q(7)+Qs bBM /Bbi`B#miBQM QMHv B7

$$\max \text{ Lip } 7_M; \text{Lip } 7_M^R \; !1 \quad :$$

6Q` 2bB/m H 6HQrb (*?2M 2i H X) - kyRN

$$\max \text{Lip } 7_M ; \text{Lip } 7_M^R \quad \max R + ; (R \quad)^{R \, G} < 1 ;$$

r?2`2 2 (y; R) Bb }t2/ GM/b i?2 MmK#2` Q7 H v2`bX

>2M+2 i?2 T`2pBQmb `2bmHi (*?2M 2i HX) 2MbmM2b 2MKMMQi7(wp)I2s BM /Bbi`B `2; `/H2bb Q7 i` BMBM; iBK2- M2m` H M2irQ`F bBx2- 2i+

AKTHB+ iBQMb 7Q` Pi?2` 6HQrb

6Q` KQbi Qi?2` max ~ Qr 7i Lip 7$_M$ ^{R} Bb mM+QMbi` (" 2B` M2 / MM 2i H
k y k)yX

>Qr2p2` - r2 + M biBHH $K(M)$ = Q $s_i$ Hv 2v ? + iBHv i?2 bm T T QMi/$s_i$ Q`2
?QK2QKQ`T?B+X

Ai b22Kb `2 bQM #H2 iQ ?QT2 7Q` #2ii2` T2`7Q`K M
+H bb b $Q(M)$ = $i s_?$ Bb i H2 T Q b i bB #H2

*QMiBMmQmbHv AM/2t2/ 6HQrb

_2+, T*QMiBMmQmbHv @*A6bM/m2 b2/ i?Q rTb`Q] +2bb

w S$_w$; l j w S$_{l j w}$(j w); s := 6(w; l);

r?2`l2 Bb +QMiBMmQmb B-M/2t p6(`;B)m#H2 MQ`K HBbBM

_2+, T *QMiBMmQmbHv @*A6b/2b2/i?Q rTb`Q +2bb

w S ; l j w S (j w); s := 6(w; l);

r?2`2 Bb +QMiBMmQmbHv BM/2t +6 (; m) #Bb2 MQ`K HBbBM

h?Bb BbQK T iB#H2 HH 2tBbiBM; MQ`K HBbBM; ~Qrb, i

6(x; m) = 72 (m) x i(m) :

r?2`2 Bb bi M/ `/ ~QrX

JmHiB@H v2` *A6b

MG@H v2` *A6 Bb Q#i BM2/ #v

$$w_y \quad S_{w_y};$$

$$l_R \quad S_{l_R; j w_y}(j w_y); \qquad w_R = 6_R(w_y; l_R);$$

$$l_G \quad S_{l_G; j w_{G\ R}}(j w_{G\ R}); \qquad s = 6_G(w_{G\ R}; l_G):$$

6B;m`2`j T?B+ H KmHiB@H v2` *A6 ;2M2` iBp2

h` BMBM; M/ BM72`2M+2

h?2 K `; BTSMBE BMi` +i #mHi2 i?2 DS, KRG Mib +HQb2/@7Q`K h?2 K `; BTMBB BMi` +i #mi2 i?2 $T_{s;\,l_{RG}}$ Q B?M ib +HQb2/@7Q`K)

:Bp2M BMi2`K2 Q/2H $[l_{RG}|s}$ - r2 + M mb2 1G"P7Q` i` BMBM;,

$$L(t) := E_{m_{RG}\ [l_{RG}|s}(\,j\,t)} \log \frac{T_{s;\,l_{RG}}(t;\ m_{RG})}{[l_{RG}|s}(m_{RG}|t)} \qquad \log T_s(t):$$

h` BMBM; M/ BM72`2M+2

h?2 K `; BTMBH BMi` +i #mHi2 i?2 DQBMi +HQb2/@7Q`K T_S, $T_{S;I_{RG}}$

:Bp2M BM72`2M+2 [$I_{RG}|S$ - r2 + M mb2 1G"P 7Q` i` BMBM;,

$$L(t) := \mathbb{E}_{m_{RG} \sim [_{I_{RG}|S}(\cdot|t)} \left[\log \frac{T_{S;I_{RG}}(t;\, m_{RG})}{[_{I_{RG}|S}(m_{RG}|t)} \right] \leq \log T_S(t):$$

i i2bi iBK2 - r2 + M mb2 $\log T_S(t)$ Ki Qi2 # Bi` `v T`2 +B B H@M K T H

Aq 12bi BK i2 K Bi R X

hQ Q#i BM{+B2MiH 72`2M+2 Q/2H t THQBi i?2 BM/2T2M/2M+2 bi`m+im`2 Q7 i?2 7Q`r `/ KQ/2H,

$$I_G \quad [_{I_G; w_G}(j w_G); \qquad w_G = s; \quad w_{G\,R} = 6_G^R(w_G; I_G);$$

$$I_R \quad [_{I_R; w_R}(j w_R); \qquad w_y = 6_R^R(w_R; I_R):$$

AM Qi?2` rQ`/b

$$[_{I_{RG}; s}(I_{RG}; s) := \prod_{`=R}^{G} [_{I_`; w}(I_`; j w):$$

AM72`2M+2 KQ/2H

hQ Q#i BM 2{+B2MiM BM 72`2M+2 KQ/2H t THHQ+Q M/ BiBQM H
BM/2T2M/2M+2 Qbi`m+i m`Q K2 i?2 7Q`r `/ KQ/2H,

$$w_G = s;$$

$$l_G \quad [l_G; w_G(j w_G); \qquad w_{G R} = 6_G^R(w_G; l_G);$$

$$l_R \quad [l_R; w_R(j w_R); \qquad w_y = 6_R^R(w_R; l_R):$$

AM Qi?2` rQ`/b

$$[l_{RG}; s(l_{RG}; s) := \prod_{` = R}^{G} [l_`; w(l_`; j w):$$

h?Bb M im b? HHv r2 B#2?i`2 2M i?2 7Q`r `/ M/ BMp2`b2 K
b K6 `2 mb2/ BM #Qi? + b2bX

AMimBiBQM

AMimBiBp2Hv- i?2 //BiBQM H$_{ij}$ $\tilde{w}$ 2HB#QrH B i*A 6 Q /2 /#

K bb i? i rQmH/ #2 KBbTH +2/ #v bBM;H2 #BD2+iB

AMimBiBQM

AMimBiBp2Hv- i?2 //BiBQM $S_{ij\tilde{w}}$ 2tB#QrH Bi*A 6 Q /2 /m#
K bb i? irQmH/ #2 KBbTH +2/ #v bBM;H2 #BD2+iBQ

S`QTQbBiBQMK +QM/BiBQMb Q-M ?2 2`2i 2`SB,bbM + ?
i? i i?2 KQs/?2Hb i?2 K2 bmTTQ`i b i?2 i `;2iX

AMimBiBQM

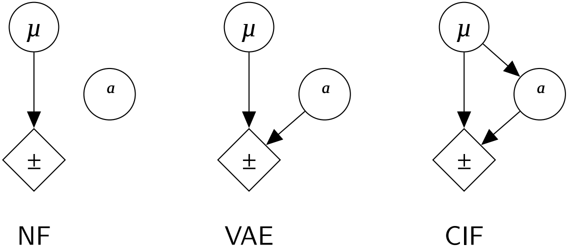
AMimBiBp2Hv- i?2 //BiBQM H$_{ij\tilde{w}}$ 2tBHBi`v Q7 i?2 //BiBQM ...
K bb i? i rQmH/ #2 KBbTH +2/ #v bBM;H2 #BD2+iBQ...

S`QTQbBiBQM, KBH/ +QM/BiBQMb QM i?2 2t2`SH$_{ij\tilde{w}}$ bbm+?
i? i i?2 KQ/2H b i?2 K2 bmTTQ`i i? i `;2iX

S`QTQbBiBQM, bm`D2+iBp2 `2 Q`2 i?2 SH$_{ij\tilde{w}}$ bbm+? i? i
K i+?2b i?2 i `;2i /Bbi`B#miBQM

Comparison with related models

CIFs may be understood as a hybrid between standard normalising flow and VAE density models:



In all cases $\pm = S(\mu;\ ^{a})$ for some family of bijections S

¦9@e0 oI ¦0 § 0§ @_§ q0} L_h0j _nju dn°0} _ @0§§0}u

	g i^~¦	G^S8 |]o·
|0 Sen° r h9ees	o:·¢V	¨:V¢V
|0 Sen° r@_Xs	o:·oP	¨:V©©
G^S]|0 Sen°	·:**k©©**	¨:¨¨**V**

Note that these ResFlows were smaller than those used by Chen et al. [2019].

Experimental Results

¦9@e0 oI ¦0 § 0§ @_§ q0} L_h0j _nju dn°0} _ @0§§0}u

	g i^~¦	G^S8 |]o·
|0 Sen° r h9ees	o:·¢V	¨:V¢V
|0 Sen° r@_Xs	o:·oP	¨:V©©
G^S]|0 Sen°	**·:k©©**	**¨:¨¨V**

Note that these ResFlows were smaller than those used by Chen et al. [2019].

We obtained similar improvements on several other problems and flow models

Experimental Results

¦9@e0 oI ¦0 § 0§ @_§ q0} L_h0j _nju dn°0} _ @0§§0}u

	g i^~¦	G^S8 |]o·
|0 Sen° r h9ees	o:·¢V	¨:V¢V
|0 Sen° r@_Xs	o:·oP	¨:V©©
G^S]|0 Sen°	·:**k©©**	**¨:¨¨V**

Note that these ResFlows were smaller than those used by Chen et al. [2019].

We obtained similar improvements on several other problems and flow models

Thank you!

S_X«}0 VI `n_j§ °n}c °_§\ 8j§\nj´ G9§0}_j_J W0n}X0 K0ȩ_X_9jj_L_ J 9jL 8}j9«L Kn«H0§

Rob Cornish, Anthony L Caterini, George Deligiannidis, and Arnaud Doucet. Relaxing bijectivity constraints with continuously-indexed normalising flows. In *^j§0}j9§_nj9e GnjT0}0jH0 nj g9H_j0 d09}j_jX*, 2020.

Tian Qi Chen, Jens Behrmann, David K Duvenaud, and Jörn-Henrik Jacobsen. Residual flows for invertible generative modeling. In *8L®9jH0 _j i0«}9e ^jTn}h9§_nj p}nH0 _jX ~´ §0h*, pages 9913–9923, 2019.

Jens Behrmann, Paul Vicol, Kuan-Chieh Wang, Roger B. Grosse, and Jörn-Henrik Jacobsen. On the invertibility of invertible neural networks, 2020.