Mini-tutorial: Normalizing flows

Zebroid Meeting

Sean Bittner

June 26, 2019

Norm flows background

- **Normalizing flows** describe the transformation of a probability density through a sequence of invertible, differentiable mappings $f : \mathcal{R}^D \rightarrow \mathcal{R}^D$.

$$z_K = f_K \circ ... \circ f_2 \circ f_1(u)$$

where $u \sim q_0(u)$ is noise.

Norm flows background

- **Normalizing flows** describe the transformation of a probability density through a sequence of invertible, differentiable mappings $f : \mathcal{R}^D \rightarrow \mathcal{R}^D$.

$$z_K = f_K \circ ... \circ f_2 \circ f_1(u)$$

  where $u \sim q_0(u)$ is noise.

- Why invertible? We use the change of variables formula (which requires invertibility) to *normalize* the density of our random variable as it flows through the sequence of such mappings.

Norm flows background

- **Normalizing flows** describe the transformation of a probability density through a sequence of invertible, differentiable mappings $f : \mathcal{R}^D \rightarrow \mathcal{R}^D$.

$$z_K = f_K \circ ... \circ f_2 \circ f_1(u)$$

  where $u \sim q_0(u)$ is noise.
- Why invertible? We use the change of variables formula (which requires invertibility) to *normalize* the density of our random variable as it flows through the sequence of such mappings.
- Why is probability density of a model useful to have?
  - density estimation – want to learn a flexible generative model of a dataset through maximum likelihood
  - variational inference – want flexible approximate posteriors (as opposed to Gaussians)

Norm flows background

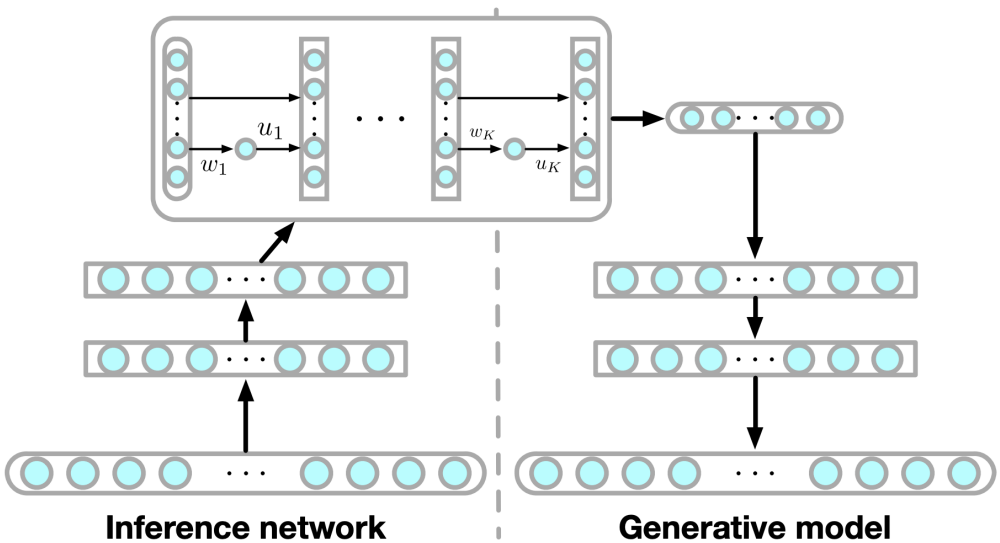
- **Normalizing flows** describe the transformation of a probability density through a sequence of invertible, differentiable mappings $f : \mathcal{R}^D \rightarrow \mathcal{R}^D$.

$$z_K = f_K \circ ... \circ f_2 \circ f_1(u)$$

  where $u \sim q_0(u)$ is noise.
- Why invertible? We use the change of variables formula (which requires invertibility) to *normalize* the density of our random variable as it flows through the sequence of such mappings.
- Why is probability density of a model useful to have?
  - density estimation – want to learn a flexible generative model of a dataset through maximum likelihood
  - variational inference – want flexible approximate posteriors (as opposed to Gaussians)
  - maximum entropy distribution approximation!

Normalizing flows 101

- What are normalizing flows in the greater context of deep generative modeling?
  - Like GANs and VAEs, normalizing flows are DGMs that generate samples in a single forward pass of a deep net (generator in GAN, decoder in VAE).
  - Unlike GANs and VAEs, normalizing flows also produce a differentiable calculation of the density of its samples along with this single forward pass.
  - Normalizing flows can be used as flexible approximate posteriors in an amortized inference setting like VAE (depicted below).



(Rezende et al. 2015)

Normalizing flows 101

- If random variable $z \sim q(z)$ is transformed via, $z' = f(z)$, then we use the **change of variables** formula to compute $q(z')$:

$$q(z') = q(f^{-1}(z')) \left| \det \frac{\partial f^{-1}(z')}{\partial z'} \right| = q(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

Normalizing flows 101

- If random variable $z \sim q(z)$ is transformed via, $z' = f(z)$, then we use the **change of variables** formula to compute $q(z')$:

$$q(z') = q(f^{-1}(z')) \left| \det \frac{\partial f^{-1}(z')}{\partial z'} \right| = q(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

- The righthand expression is sometimes faster, and useful in settings where we have access to z
  - like in variational inference,
  - but not density estimation.

Normalizing flows 101

NF generative model

$$z_K = f_K \circ ... \circ f_2 \circ f_1(u), \; u \sim q_0(u)$$

+ change of variables formulas

$$q(z') = q(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

$$\Longrightarrow \textbf{ NF density calculation}$$

$$q_K(z_K) = q_0(z_0) \prod_{k=1}^{K} \left| \det \frac{\partial f_k(z_k)}{\partial z_{k-1}} \right|^{-1}$$

- In most settings, we're optimizing the log density.

Log density

$$\log q_K(z_K) = \log q_0(z_0) - \sum_{k=1}^{K} \log \left| \det \frac{\partial f_k(z_k)}{\partial z_{k-1}} \right|$$

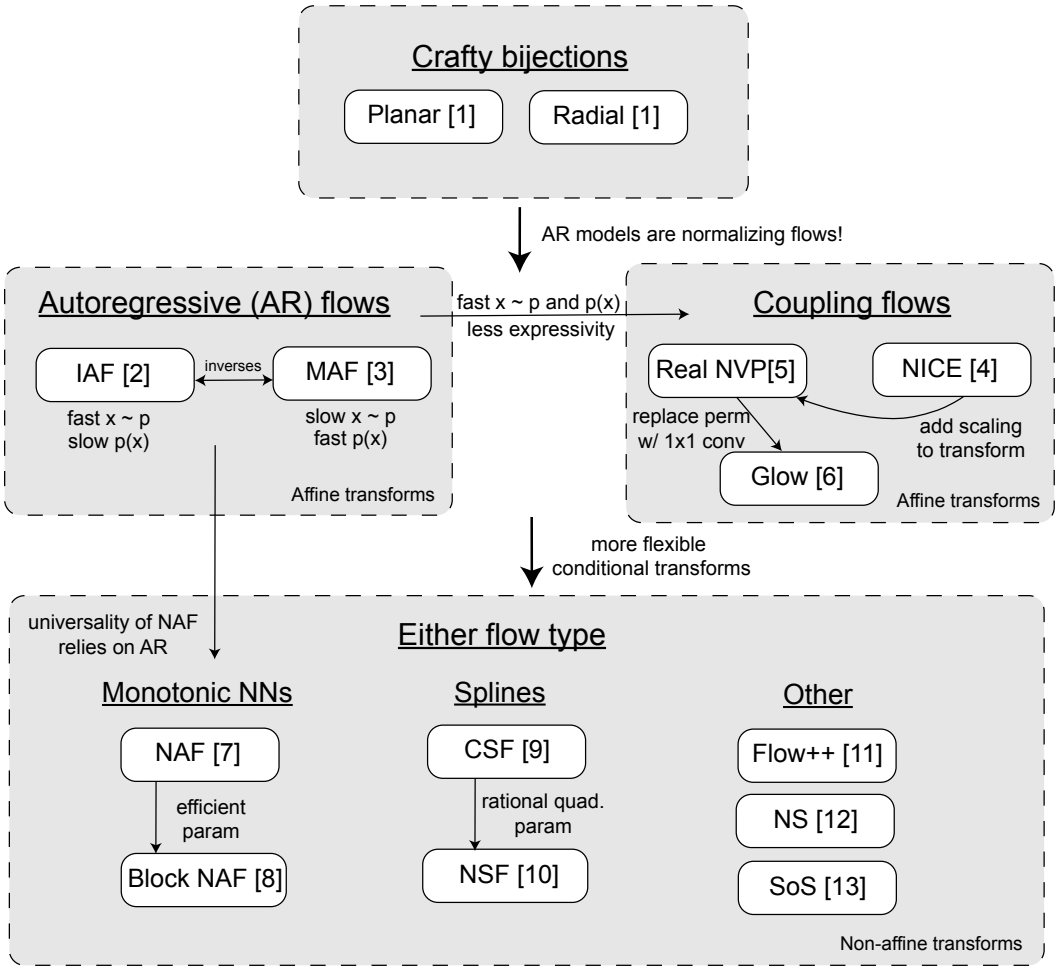
Normalizing flows 101

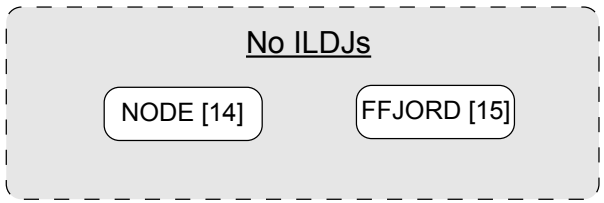
- Current research on normalizing flows focuses on designing function classes for f that are
  - expressive
  - have fast $\log \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$ or "[inverse] log det jacs" or "[I]LDJs"

Log density

$$\log q_K(z_K) = \log q_0(z_0) - \sum_{k=1}^{K} \log \left| \det \frac{\partial f_k(z_k)}{\partial z_{k-1}} \right|$$

Normalizing flows





[1] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[2] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[4] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[7] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*, 2018.

[8] Nicola De Cao, Ivan Titov, and Wilker Aziz. Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*, 2019.

[9] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. *arXiv preprint arXiv:1906.02145*, 2019.

[10] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[11] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. *arXiv preprint arXiv:1905.02325*, 2019.

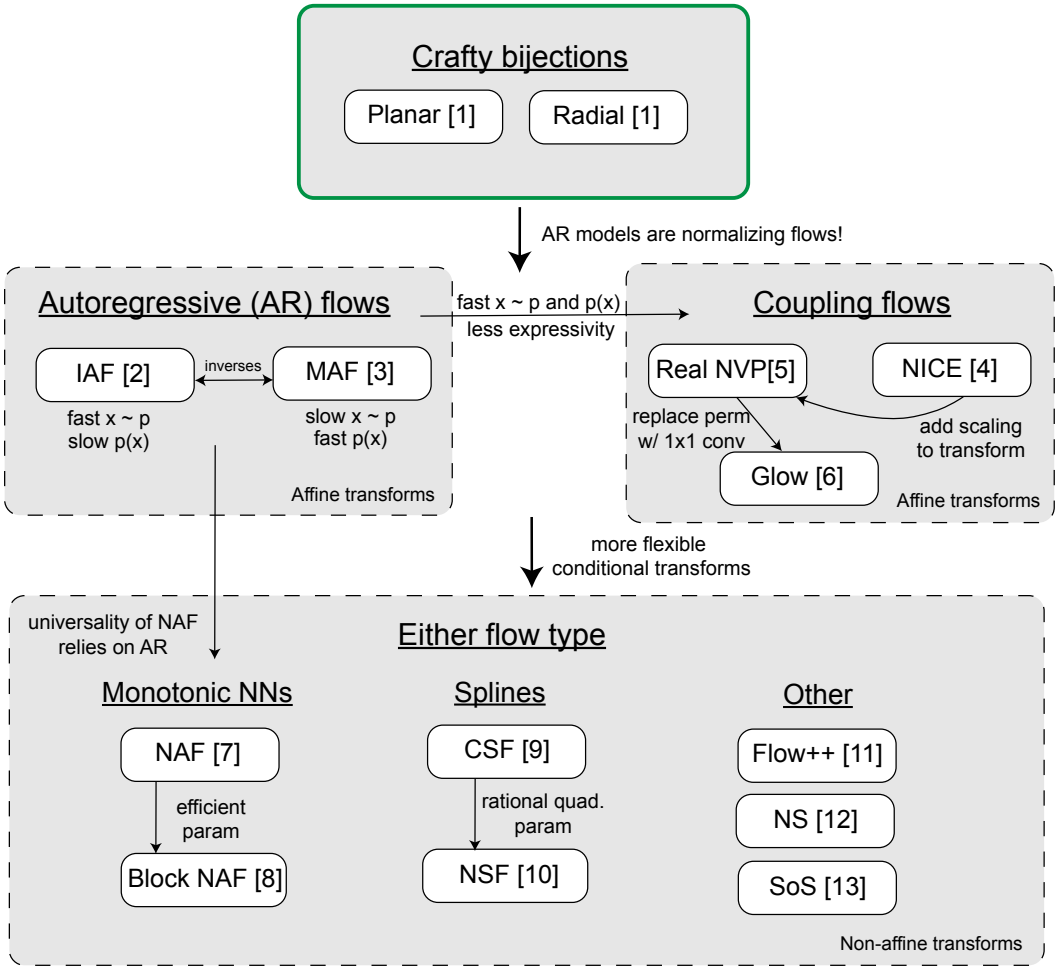
[12] Zachary M Ziegler and Alexander M Rush. Latent normalizing flows for discrete sequences. *arXiv preprint arXiv:1901.10548*, 2019.

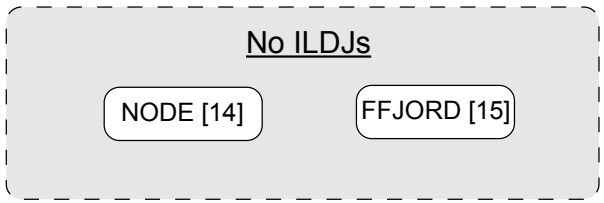
[13] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[14] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[15] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

Normalizing flows roadmap





[1] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[2] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[4] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[7] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*, 2018.

[8] Nicola De Cao, Ivan Titov, and Wilker Aziz. Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*, 2019.

[9] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. *arXiv preprint arXiv:1906.02145*, 2019.

[10] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[11] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. *arXiv preprint arXiv:1905.02325*, 2019.

[12] Zachary M Ziegler and Alexander M Rush. Latent normalizing flows for discrete sequences. *arXiv preprint arXiv:1901.10548*, 2019.

[13] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[14] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[15] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

Crafty bijections

Planar flows

$$f(\mathbf{z}) = \mathbf{z} + \mathbf{v}h(\mathbf{w}^\top \mathbf{z} + b)$$

$$\psi(\mathbf{z}) = h'(\mathbf{w}^\top \mathbf{z} + b)\mathbf{w}$$

$$\left|\det \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}}\right| = \left|1 + \mathbf{v}^\top \psi(\mathbf{z})\right|$$

Radial flows

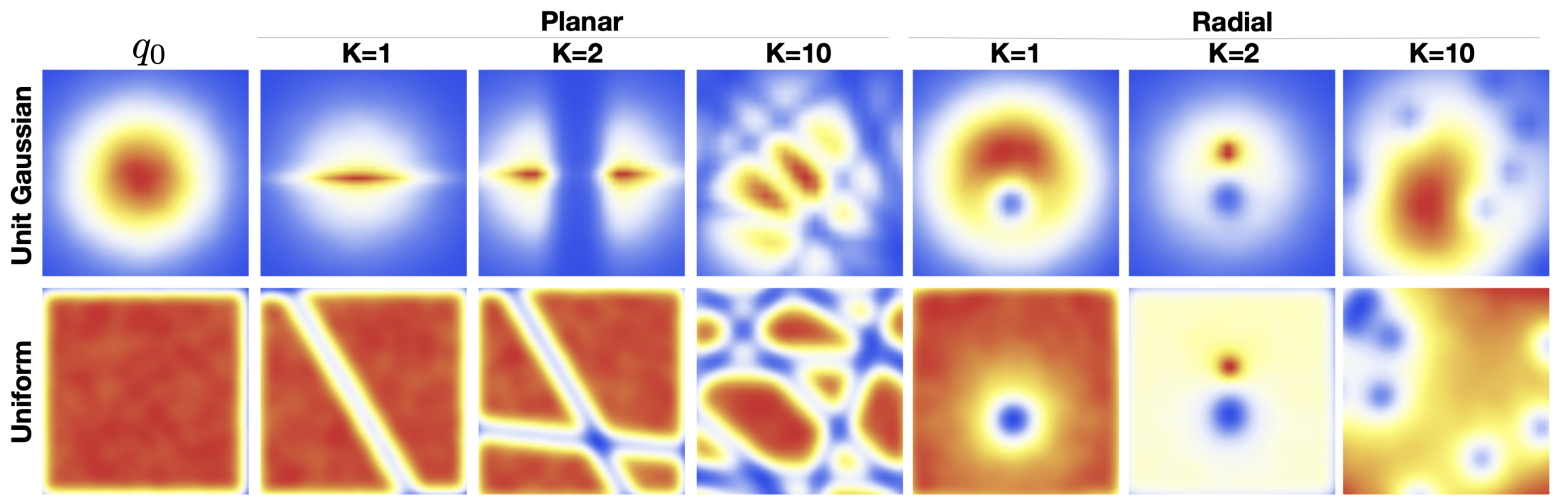
$$f(\mathbf{z}) = \beta h(\alpha, r)(\mathbf{z} - \mathbf{z}_0)$$

$$r = |\mathbf{z} - \mathbf{z}_0|\,, \;\; h(\alpha, r) = \frac{1}{\alpha + r}$$

$$\left|\det \tfrac{\partial f(\mathbf{z})}{\partial \mathbf{z}}\right| = [1 + \beta h(\alpha, r)]^{d-1}\,[1 + \beta h(\alpha, r) + \beta h'(\alpha, r) r]$$

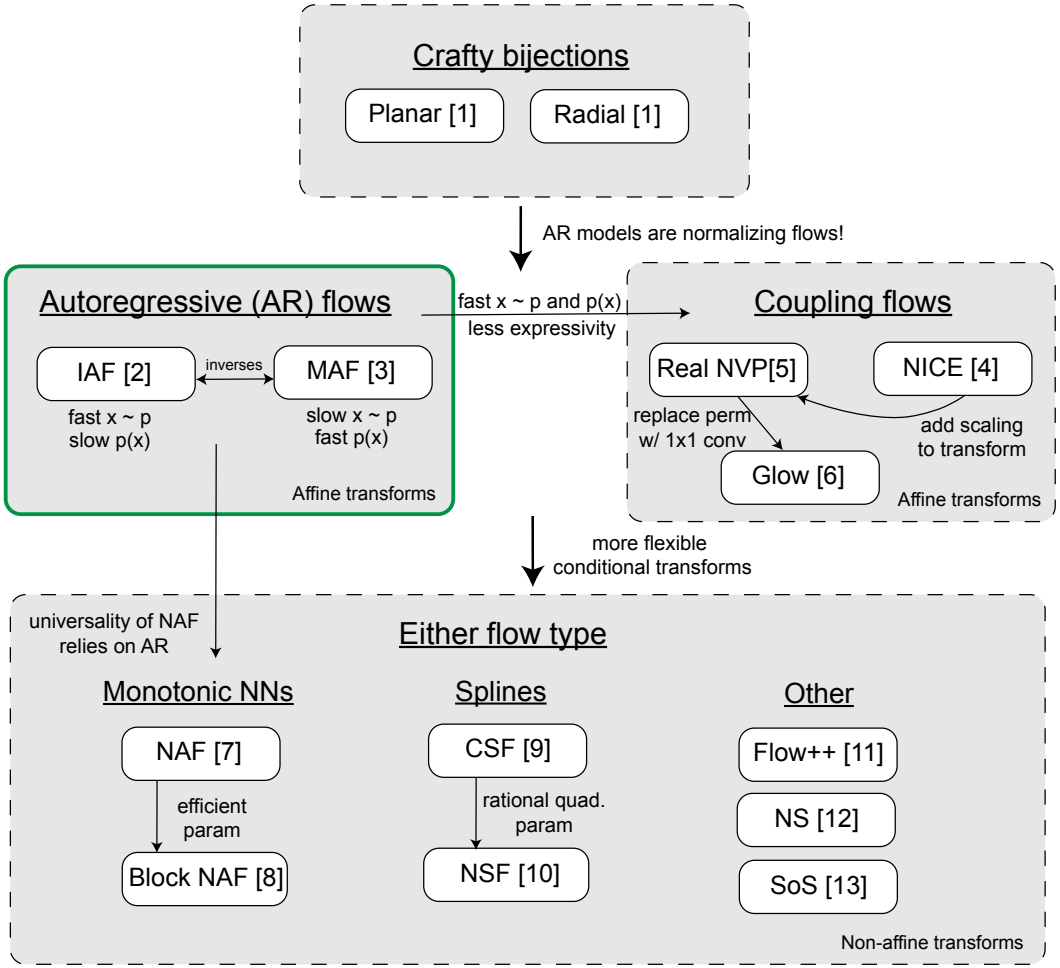
$$\mathbf{z}_K = f_K \circ \ldots \circ f_1(\mathbf{u})$$

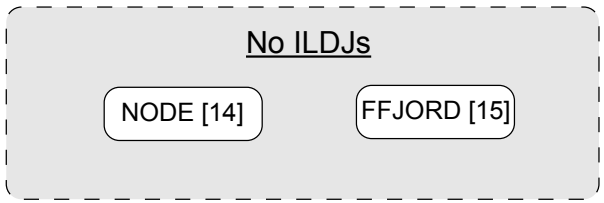
$$\log q_K(\mathbf{z}_K) = \log q_0(\mathbf{u}) - \sum_{k=1}^{K} \left|\det \frac{\partial f(\mathbf{z}_k)}{\partial \mathbf{z}_k}\right|$$



(Rezende et al. 2015)

Normalizing flows roadmap





[1] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[2] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[4] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[7] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*, 2018.

[8] Nicola De Cao, Ivan Titov, and Wilker Aziz. Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*, 2019.

[9] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. *arXiv preprint arXiv:1906.02145*, 2019.

[10] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[11] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. *arXiv preprint arXiv:1905.02325*, 2019.

[12] Zachary M Ziegler and Alexander M Rush. Latent normalizing flows for discrete sequences. *arXiv preprint arXiv:1901.10548*, 2019.

[13] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[14] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[15] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

Autoregressive flows

- Autoregressive models such as MADE, NADE, PixelCNN, PixelRNN, Wavenet have been very successful generative models.

$$p(x) = \prod_{i=1}^{D} p(x_i \mid x_{1:i-1})$$

$$p(x_i \mid x_{1:i-1}) = \mathcal{N}(x_i \mid \mu_i, (\exp \alpha_i)^2), \ \mu_i = f_{\mu_i}(x_{1:i-1}), \ \alpha_i = f_{\alpha_i}(x_{1:i-1})$$

Autoregressive flows

- Autoregressive models such as MADE, NADE, PixelCNN, PixelRNN, Wavenet have been very successful generative models.

$$p(x) = \prod_{i=1}^{D} p(x_i \mid x_{1:i-1})$$

$$p(x_i \mid x_{1:i-1}) = \mathcal{N}(x_i \mid \mu_i, (\exp \alpha_i)^2), \;\; \mu_i = f_{\mu_i}(x_{1:i-1}), \;\; \alpha_i = f_{\alpha_i}(x_{1:i-1})$$

- **Inverse autoregressive flows (IAF)** - (Kingma et al. 2016):
  - Key insight: AR models with Gaussian conditional distributions are normalizing flows! Can be viewed as an affine transformation of a standard normal $u \sim \mathcal{N}(0, I)$, yielding a lower-traingular Jacobian.

$$x_i = u_i \exp \alpha_i + \mu_i$$

$$x = g(u) \qquad \log \left| \det \frac{\partial g(u)}{\partial u} \right| = \sum_{i=1}^{D} \log \alpha_i$$

Autoregressive flows

- Autoregressive models such as MADE, NADE, PixelCNN, PixelRNN, Wavenet have been very successful generative models.

$$p(x) = \prod_{i=1}^{D} p(x_i \mid x_{1:i-1})$$

$$p(x_i \mid x_{1:i-1}) = \mathcal{N}(x_i \mid \mu_i, (\exp \alpha_i)^2), \;\; \mu_i = f_{\mu_i}(x_{1:i-1}), \;\; \alpha_i = f_{\alpha_i}(x_{1:i-1})$$

- **Inverse autoregressive flows (IAF)** - (Kingma et al. 2016):
  - Key insight: AR models with Gaussian conditional distributions are normalizing flows! Can be viewed as an affine transformation of a standard normal $u \sim \mathcal{N}(0, I)$, yielding a lower-traingular Jacobian.

$$x_i = u_i \exp \alpha_i + \mu_i$$

$$x = g(u) \qquad \log \left| \det \frac{\partial g(u)}{\partial u} \right| = \sum_{i=1}^{D} \log \alpha_i$$

  - We can still have a normalizing flow with whatever function we desire for f_{μ_i} and f_{α_i}.

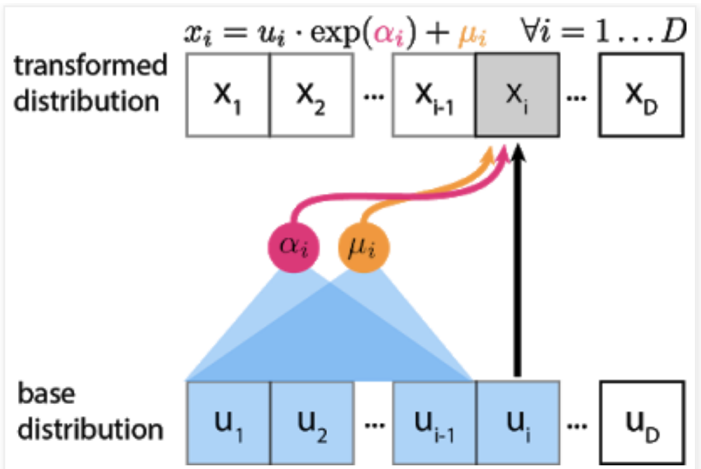
Autoregressive flows

- **Inverse autoregressive flows (IAF)** - (Kingma et al. 2016):

$$x_i = u_i \exp \alpha_i + \mu_i$$

$$\mu_i = f_{\mu_i}(u_{1:i-1}), \qquad \alpha_i = f_{\alpha_i}(u_{1:i-1})$$

- Fast $x \sim p(x)$
- Slow $p(x)$



https://blog.evjang.com/2018/01/nf2.html

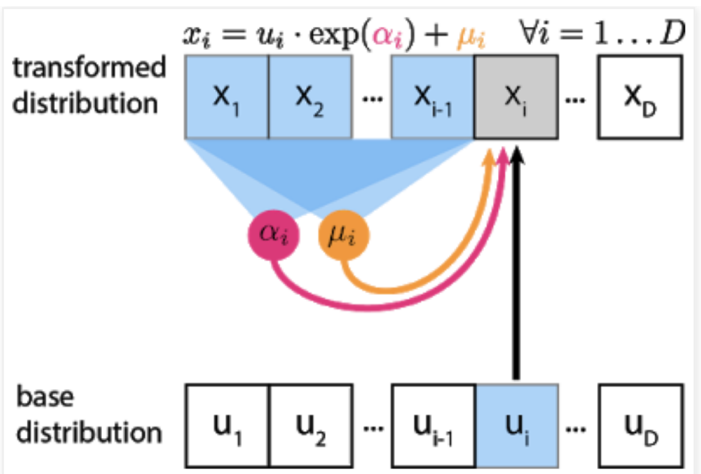
Autoregressive flows

- **Masked autoregressive flows (MAF)** - (Papamakarios et al. 2017):

$$x_i = u_i \exp \alpha_i + \mu_i$$

$$\mu_i = f_{\mu_i}(x_{1:i-1}), \qquad \alpha_i = f_{\alpha_i}(x_{1:i-1})$$

- Slow $x \sim p(x)$
- Fast $p(x)$

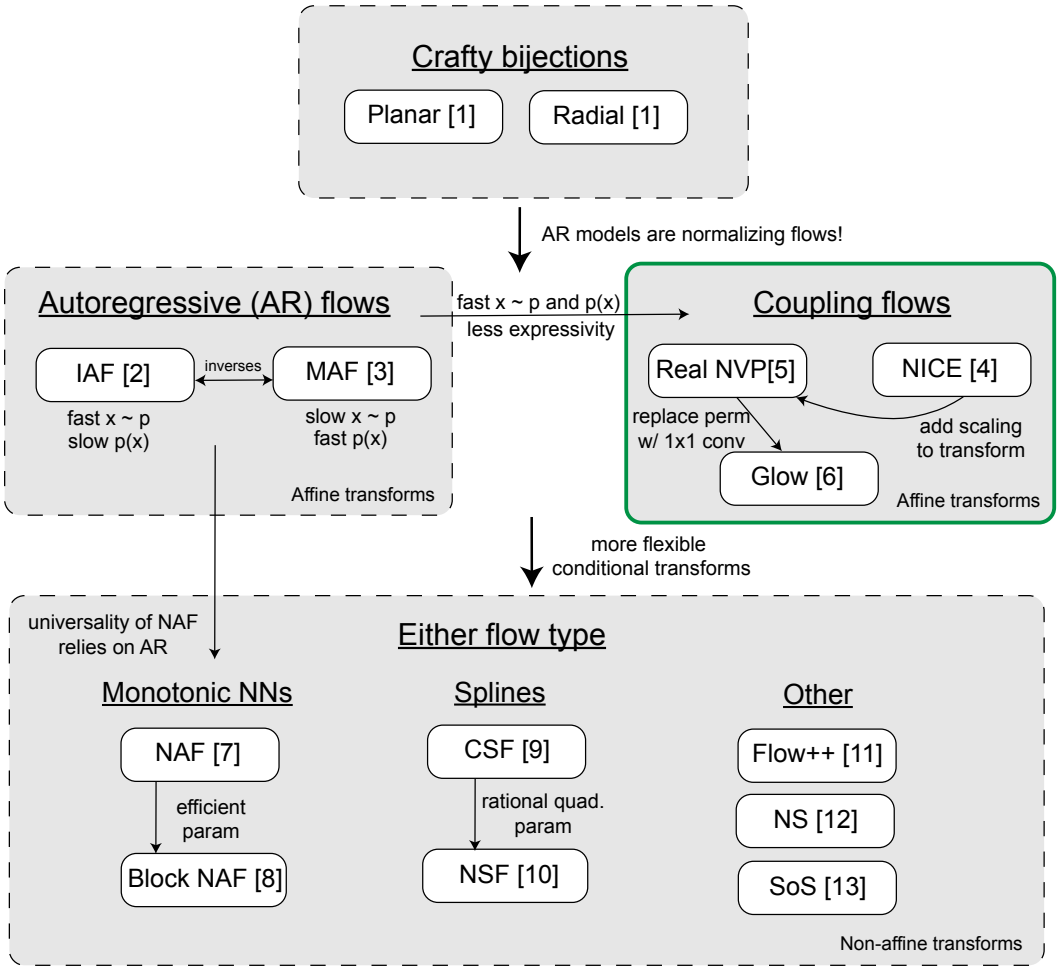


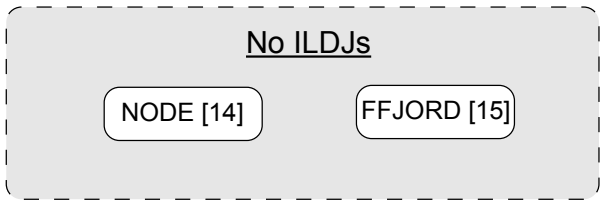
https://blog.evjang.com/2018/01/nf2.html

Autoregressive flows

- How to get AR normalizing flows to work:
  - Such a factorization is a pretty strong inductive bias.
  - Most AR normalizing flows stack multiple AR factorizations on top of each other with different orderings.
  - This is achieved by randomly permuting the elements between AR stacks.

Normalizing flows roadmap





[1] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[2] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[4] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[7] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*, 2018.

[8] Nicola De Cao, Ivan Titov, and Wilker Aziz. Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*, 2019.

[9] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. *arXiv preprint arXiv:1906.02145*, 2019.

[10] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[11] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. *arXiv preprint arXiv:1905.02325*, 2019.

[12] Zachary M Ziegler and Alexander M Rush. Latent normalizing flows for discrete sequences. *arXiv preprint arXiv:1901.10548*, 2019.

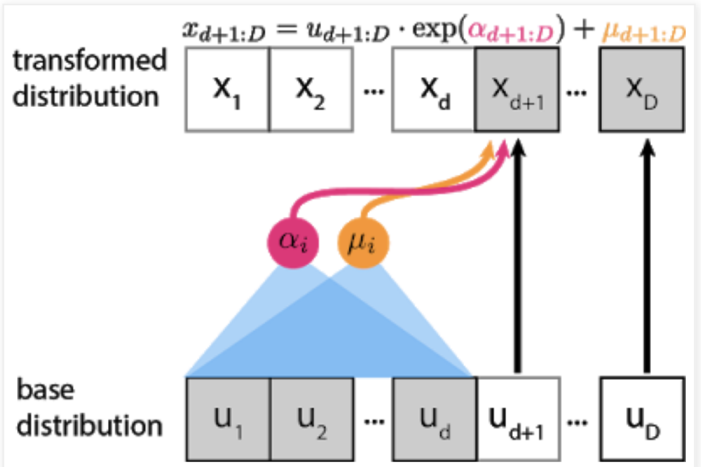
[13] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[14] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[15] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

Coupling flows

- **Real NVP** - (Dinh et al. 2017):
- **Coupling transforms** (rather than autoregressive transforms) condition the latter $D - d$ samples on the first d.



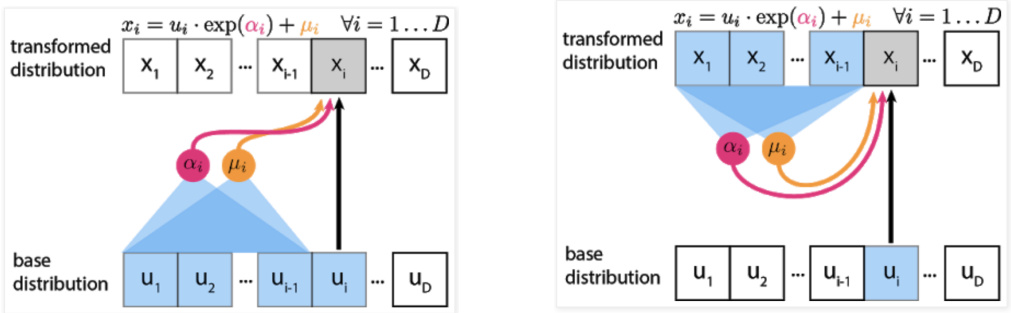
https://blog.evjang.com/2018/01/nf2.html

- Doing this, we get fast $x \sim p(x)$ AND $p(x)$.
- Creating a stack of such couplings increases expressivity. Permutations are still done between stacks.
- Side note: NICE (Dinh et al. 2014), which is oft-cited, is real-NVP with $\alpha_i = 1$

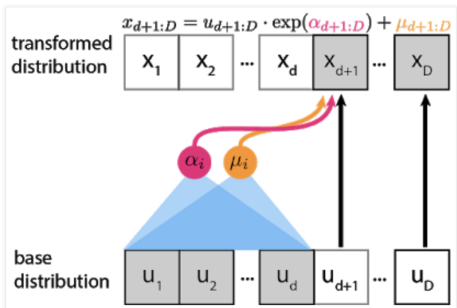
AR vs Coupling

- Autoregressive: expressive, but slow in either sampling or density calc.
- Coupling: less expressive, but fast in both directions.

Autoregressive



Coupling



Glow

- Real NVP, but rather than permute at each layer apply an invertible matrix transformation

$$W = PL(U + diag(\mathbf{s}))$$

- P is permuation matrix, L is lower triangular ones, U is upper triangular with zeros on diagonal, $s_i \in \mathcal{R}_+$.

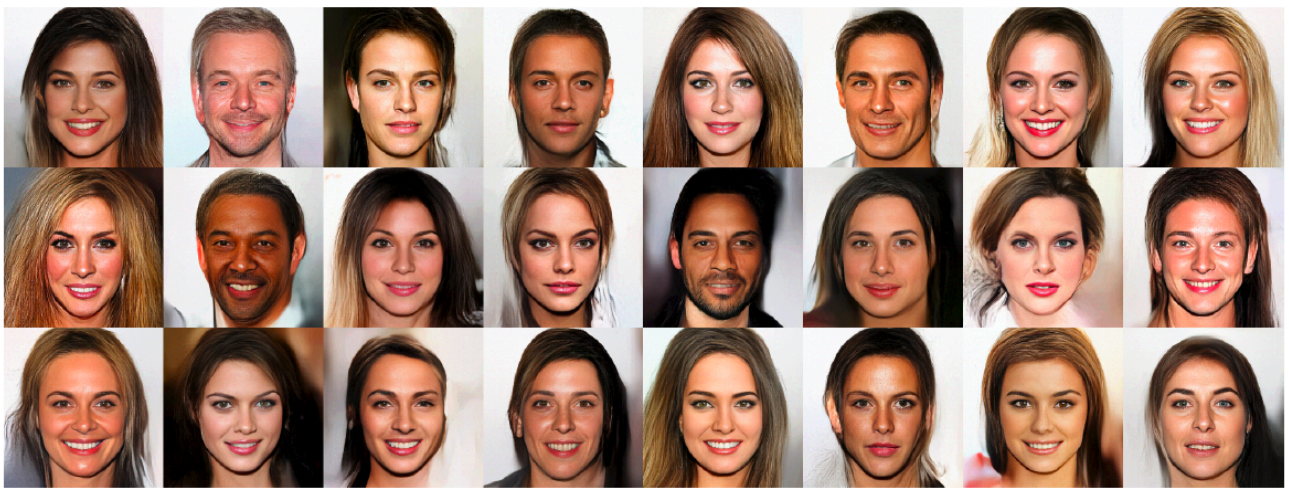
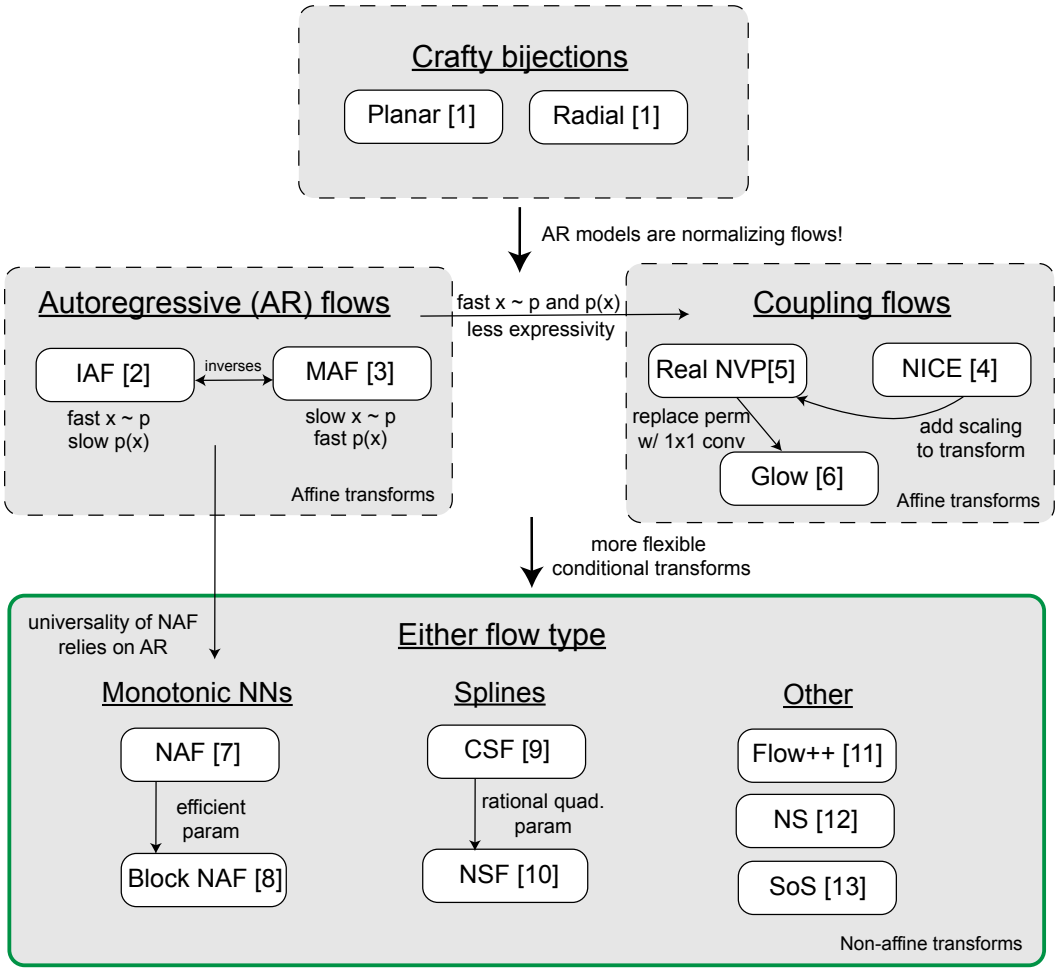
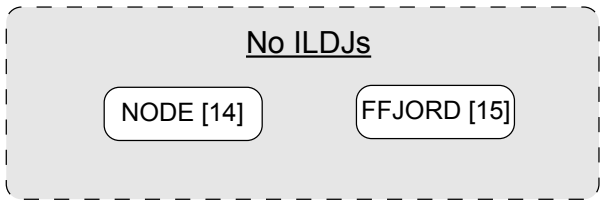


Figure 4: Random samples from the model, with temperature 0.7.

(Kingma et al. 2018)

Normalizing flows roadmap





[1] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[2] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[4] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[7] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*, 2018.

[8] Nicola De Cao, Ivan Titov, and Wilker Aziz. Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*, 2019.

[9] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. *arXiv preprint arXiv:1906.02145*, 2019.

[10] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[11] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. *arXiv preprint arXiv:1905.02325*, 2019.

[12] Zachary M Ziegler and Alexander M Rush. Latent normalizing flows for discrete sequences. *arXiv preprint arXiv:1901.10548*, 2019.

[13] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[14] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[15] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

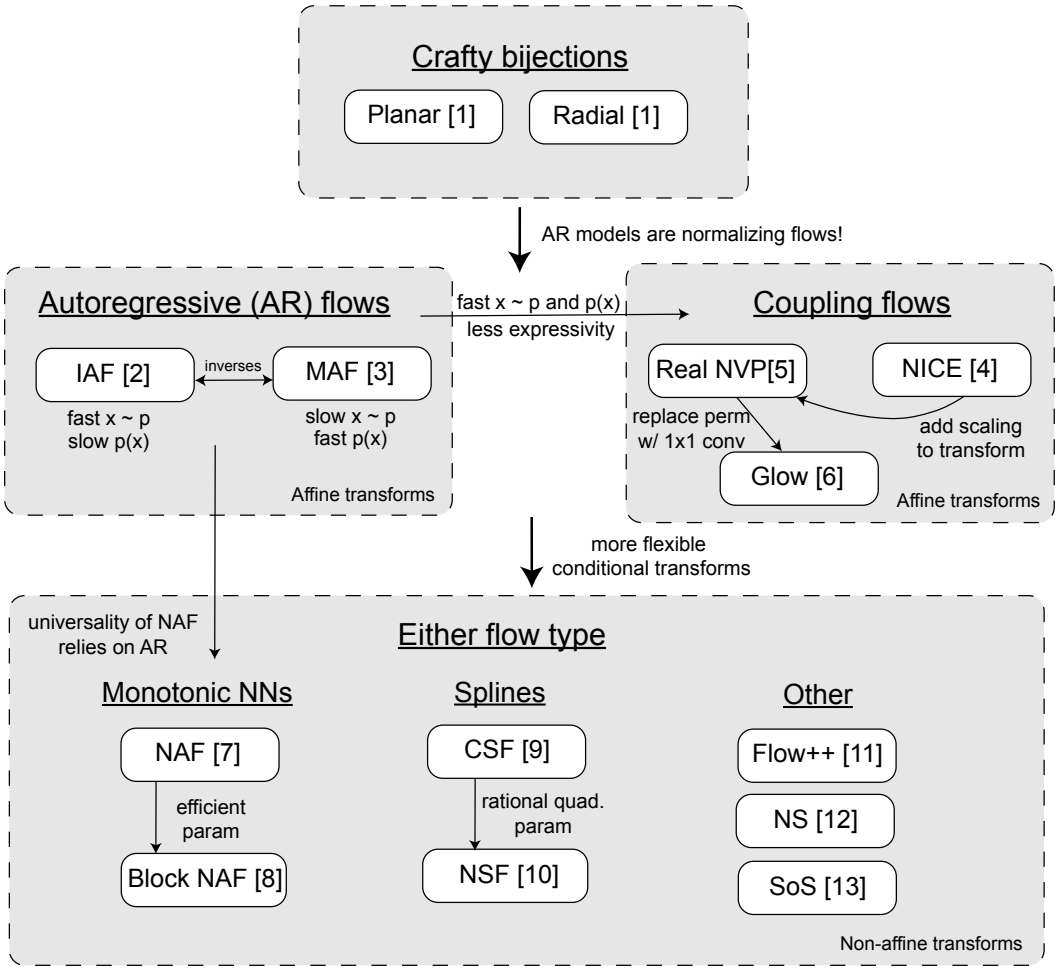
Non-affine transforms

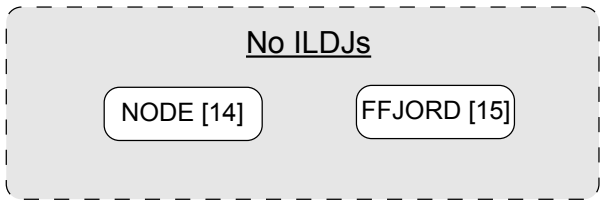
Table 1: Test log likelihood (in nats) for UCI datasets and BSDS300, with error bars corresponding to two standard deviations. NAF†, Block-NAF†, and SOS† report error bars across repeated runs rather than across the test set. FFJORD do not report error bars. Superscript* indicates results are taken from the existing literature.

Model	POWER	GAS	HEPMASS	MINIBOONE	BSDS300
FFJORD*	0.46	8.59	-14.92	-10.43	157.40
Glow	0.42 ± 0.01	12.24 ± 0.03	-16.99 ± 0.02	-10.55 ± 0.45	156.95 ± 0.28
Q-NSF (C)	0.64 ± 0.01	12.80 ± 0.02	-15.35 ± 0.02	-9.35 ± 0.44	157.65 ± 0.28
RQ-NSF (C)	0.64 ± 0.01	13.09 ± 0.02	-14.75 ± 0.03	-9.67 ± 0.47	157.54 ± 0.28
MAF	0.45 ± 0.01	12.35 ± 0.02	-17.03 ± 0.02	-10.92 ± 0.46	156.95 ± 0.28
Q-NSF (AR)	0.66 ± 0.01	12.91 ± 0.02	-14.67 ± 0.03	-9.72 ± 0.47	157.42 ± 0.28
NAF*†	0.62 ± 0.01	11.96 ± 0.33	-15.09 ± 0.40	-8.86 ± 0.15	157.73 ± 0.04
Block-NAF*†	0.61 ± 0.01	12.06 ± 0.09	-14.71 ± 0.38	-8.95 ± 0.07	157.36 ± 0.03
SOS*†	0.60 ± 0.01	11.99 ± 0.41	-15.15 ± 0.10	-8.90 ± 0.11	157.48 ± 0.41
RQ-NSF (AR)	0.66 ± 0.01	13.09 ± 0.02	-14.01 ± 0.03	-9.22 ± 0.48	157.31 ± 0.28

(Durkan et al. 2019)

Normalizing flows roadmap





[1] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[2] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[4] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[7] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*, 2018.

[8] Nicola De Cao, Ivan Titov, and Wilker Aziz. Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*, 2019.

[9] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. *arXiv preprint arXiv:1906.02145*, 2019.

[10] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[11] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. *arXiv preprint arXiv:1905.02325*, 2019.

[12] Zachary M Ziegler and Alexander M Rush. Latent normalizing flows for discrete sequences. *arXiv preprint arXiv:1901.10548*, 2019.

[13] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[14] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[15] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

[1] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[2] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[4] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[7] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. *arXiv preprint arXiv:1804.00779*, 2018.

[8] Nicola De Cao, Ivan Titov, and Wilker Aziz. Block neural autoregressive flow. *arXiv preprint arXiv:1904.04676*, 2019.

[9] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. *arXiv preprint arXiv:1906.02145*, 2019.

[10] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[11] Priyank Jaini, Kira A Selby, and Yaoliang Yu. Sum-of-squares polynomial flow. *arXiv preprint arXiv:1905.02325*, 2019.

[12] Zachary M Ziegler and Alexander M Rush. Latent normalizing flows for discrete sequences. *arXiv preprint arXiv:1901.10548*, 2019.

[13] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[14] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[15] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.