Invertible Convolutional Flow

M. Karami, J. Sohl-Dickstein, D. Schuurmans, L. Dinh, D. Duckworth

University of Alberta, Google
Two ways to improve expressivity of normalizing flow:

- Invertible convolution filter
- Invertible nonlinear gates
Circular Convolution

- Linear convolution of two sequences when one is padded cyclically
  \[ y(i) := \sum_{n=0}^{N-1} x(n) w(i - n) \mod N \]
- Jacobian of this convolution forms a circulant matrix
- Its eigenvalues are equal to the DFT of \( w \), so
  \[ \log |\det J_y| = \sum_{n=0}^{N-1} \log |w_F(n)| \]
- The circular convolution-multiplication property
  \[ y_F(k) = w_F(k) x_F(k) \]
- Inverse operation (deconvolution)
  \[ x_F(n) = w_F^{-1}(n) y_F(n) \]
- These can be evaluated in \( O(N \log N) \) time in the frequency domain, using FFT algorithms.
Symmetric Convolution

- Using even-symmetric expansion
  \[ \hat{x}(n) = \varepsilon\{x(n)\} := \begin{cases} 
  x(n) & n = 0, 1, \ldots, N - 1 \\
  x(-n - 1) & n = -N, \ldots, -1 
\end{cases} \]

- The symmetric convolution can be defined as
  \[ y = w *_s x := \mathcal{R}\{\hat{x} \odot \hat{w}\} \]

- The convolution-multiplication property holds for DCT of operands
  \[ \mathcal{F}_{dct}\{y\} = \mathcal{F}_{dct}\{w\} \odot \mathcal{F}_{dct}\{x\} \]

- The convolution, its Jacobian-determinant and inversion (deconvolution) can be performed efficiently in \(O(N \log N)\).
Let $x_1$ and $x_2$ are the disjoint parts of the input $x$.

A **data-adaptive convolution** is defined by convolving $x_2$ with an arbitrary function of $x_1$

$$f_*(x_2; x_1) = w(x_1) * x_2$$

Using any of the invertible convolutions, this transform is invertible with cheap inversion and cheap log-det-Jacobian computation.
Pointwise nonlinear bijectors

- log-det-Jacobian term in the log-likelihood equation can be interpreted as a regularizer.
- If we would like to encourage some desirable statistical properties, formulated by a regularizer $\psi(y)$, in intermediate layers of a flow-based model, we can do so by carefully designing nonlinearities $y = f(x)$.

- $f(x)$ is obtained by solving the differential equation
  \[ \left| \frac{\partial f^{-1}}{\partial y} \right| = \frac{\partial g}{\partial y} = e^{\gamma(y)} \]

- For $l_1$ regularization, inducing sparsity, this leads to the **S-Log gate** defined as
  \[ f_\alpha(x) = \frac{\text{sign}(x)}{\alpha} \ln(\alpha|x| + 1) \]
  \[ f^{-1}_\alpha(y) = \frac{\text{sign}(y)}{\alpha} (e^{\alpha|y|} - 1) \]
Combining the invertible convolution, element-wise multiplication and nonlinear bijectors, we achieve a more expressive flow in the coupling form:

\[
\begin{align*}
    y_1 &= x_1 \\
    y_2 &= f_\alpha' \left( s(x_1) \odot f_\alpha(w(x_1) \ast x_2) \right) + t(x_1)
\end{align*}
\]