Input Convex Gradient Networks

Jack Richter-Powell[1], Jonathan Lorraine[3,4], Brandon Amos[2]

McGill University[1], FAIR[2], University of Toronto[3], Vector Institute[4]

Motivation

- Kantorovich duality relates Optimal Transport primal to dual over convex functions for squared-Euclidean cost [7].
- From Brenier's theorem[2], we know Optimal Transport map T^* exists and is realized as the gradient of the optimal dual potential $T^* = (\nabla\varphi^*)$.
- This tells us parameterizing convex gradients allows us to approximate OT map between two densities. Has been explored in [4], [5] and for density estimation in [3].

Existing Method

- Existing method for modeling optimal transport maps involves using an Input Convex Neural Network [1].
- The network models a scalar potential $N_\theta : \mathbb{R}^n \to \mathbb{R}$ which is then differentiated to produce ∇N_θ.

ICGN - High Level Details

- Our approach avoids differentiation of a potential – which can causes numerical issues.
- Instead, we numerically compute a line integral of a symmeteric PSD 2-tensor $[Df]^T Df$ where $f : \mathbb{R}^n \to \mathbb{R}^m$.
- We use autograd to compute JVPs and VJPs without explicitly constructing matrices in the matrix-vector products.
- We can use any quadrature method for the integral.

Motivating Theorem

Theorem 2:
The Jacobian of N_θ, DN_θ takes the form

$$DN_\theta = [DM_\theta]^T DM_\theta$$

since this matrix is symmetric PSD, this implies there exists a convex $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $N_\theta = \nabla\varphi$ (see [6]). So our model paramterizes a convex gradient.

Definition of the model

For a Neural Network $M_\theta : \mathbb{R}^n \to \mathbb{R}^m$ that satisfies the PDE

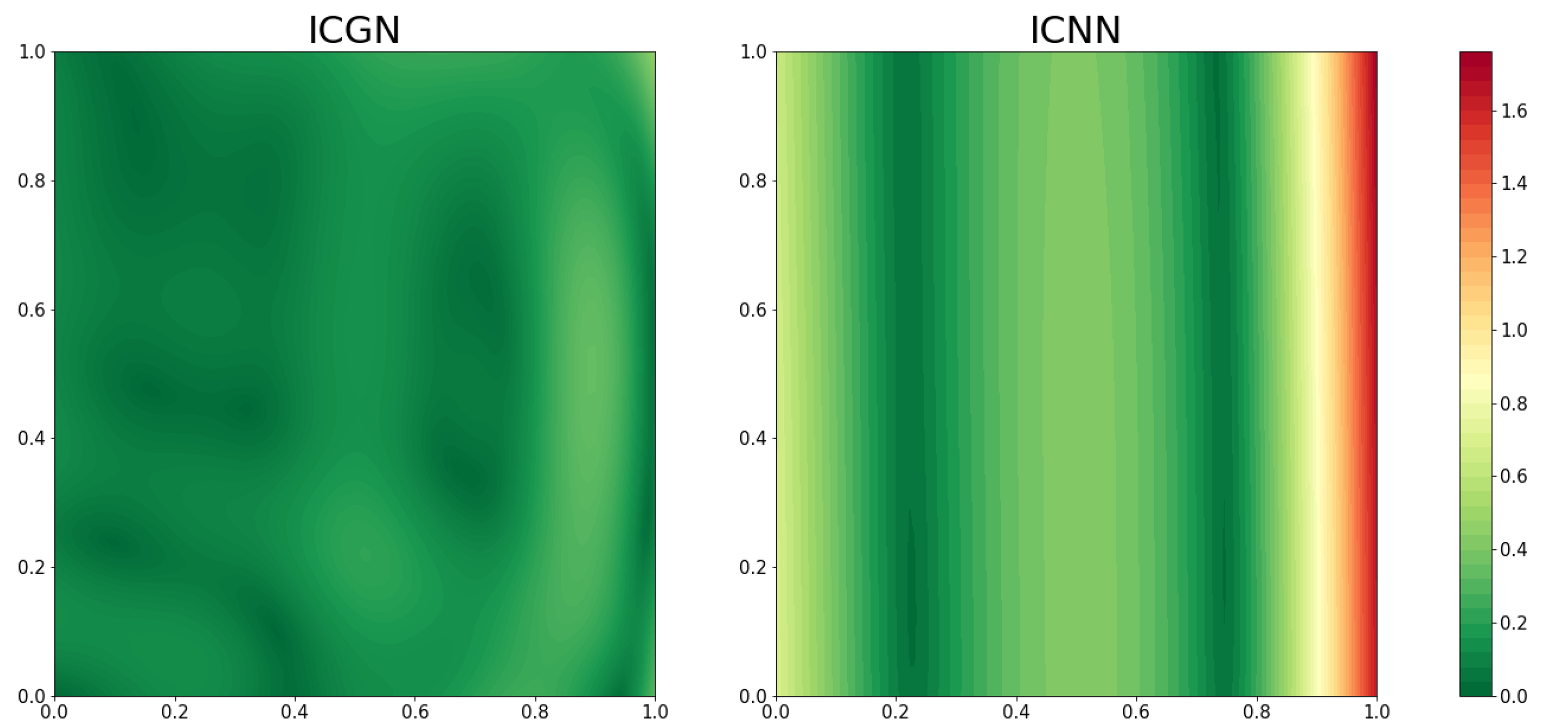
$$\frac{\partial^2 M_\theta}{\partial x^k \partial x^i} \cdot \frac{\partial M_\theta}{\partial x^j} = \frac{\partial^2 M_\theta}{\partial x^k \partial x^j} \cdot \frac{\partial M_\theta}{\partial x^i} \qquad \forall 1 \le i, j, k \le n \tag{1}$$

we define $N_\theta : \mathbb{R}^n \to \mathbb{R}^n$, an Input Convex Gradient Network as

$$N_\theta(x) = \int_0^1 [D(M_\theta)_{sx}]^T D(M_\theta)_{sx} x ds$$

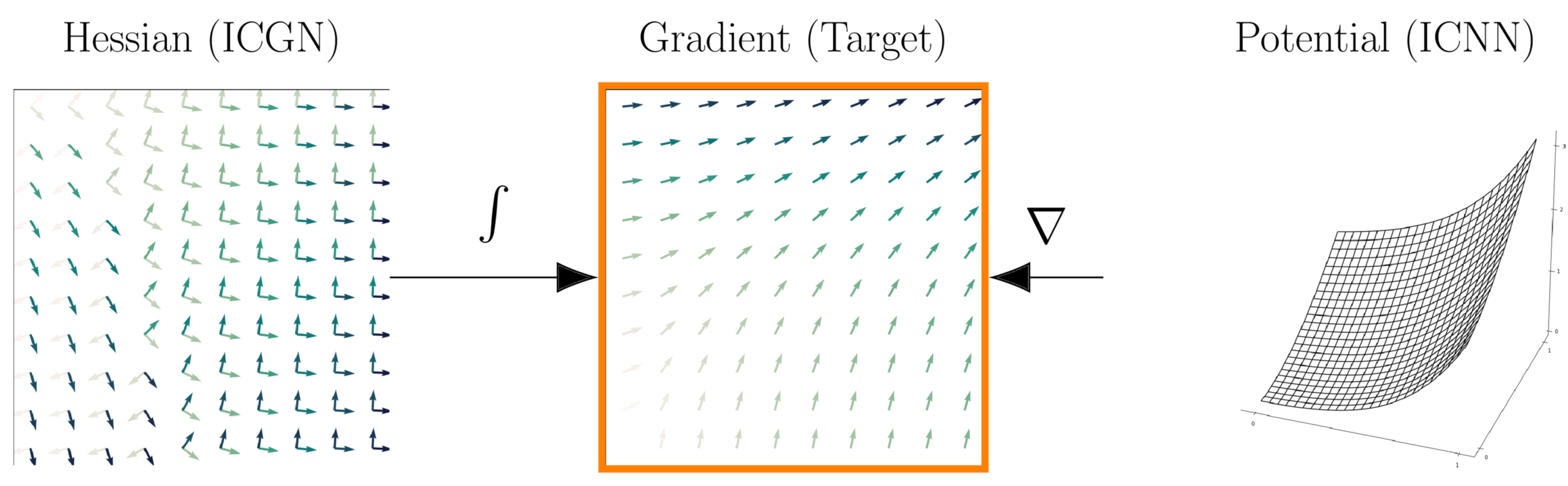
Experiment 1 - Fitting a Toy Potential Field

We compared the gradient of a 1-layer ICNN to a 1-layer ICGN for fitting a target map $T : \mathbb{R}^2 \to \mathbb{R}^2$.



We show the squared Euclidean error when fitting the target T with either the gradient of the ICNN potential or ICGN approach.

Theoretical comparison: ICGN vs ICNN gradients



Visualization of modeling the target gradient as either the ICNN potential's gradient or ICGN model, which involves the integration of the Hessian.

Experiment 1 Details

The target map is:

$$T(x, y) = \begin{pmatrix} 4x^3 + \frac{1}{2}y + x \\ 3y - y^2 + \frac{1}{2}x \end{pmatrix}$$

This is the gradient of a convex polynomial on $[0, 1]^2$.

Size of models used

Model	Layers	Hidden	Params (Total)
ICNN	1	25	78
ICGN	1	5	15

Takeaway:
The ICGN is able to fit the target T better with far fewer parameters.

Limitations / Next Steps

- Current framework can only handle one layer networks M_θ, due to PDE (1) constraint.
- Interesting future work is generalizing to deeper networks.
- We deal with potential fields on $\mathbb{R}^n$, but similar methods could be applied to model exact 1-forms on Riemannian Manifolds. Brenier's theorem has already seen extensions here[7].

References

[1] Brandon Amos, Lei Xu, and J Zico Kolter. Input convex neural networks. In *International Conference on Machine Learning*, pages 146–155. PMLR, 2017.

[2] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[3] Chin-Wei Huang, Ricky T. Q. Chen, Christos Tsirigotis, and Aaron C. Courville. Convex potential flows: Universal probability distributions with optimal transport and convex optimization. *CoRR*, abs/2012.05942, 2020. URL https://arxiv.org/abs/2012.05942.

[4] Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pages 6672–6681. PMLR, 2020.

[5] Petr Mokrov, Alexander Korotin, Lingxiao Li, Aude Genevay, Justin Solomon, and Evgeny Burnaev. Large-scale wasserstein gradient flows, 2021.

[6] Jack Richter-Powell, Jonathan Lorraine, and Brandon Amos. Input convex gradient networks, 2021.

[7] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003. ISBN 9780821833124.