Harpoon: Generalised Manifold Guidance for Conditional Tabular Diffusion

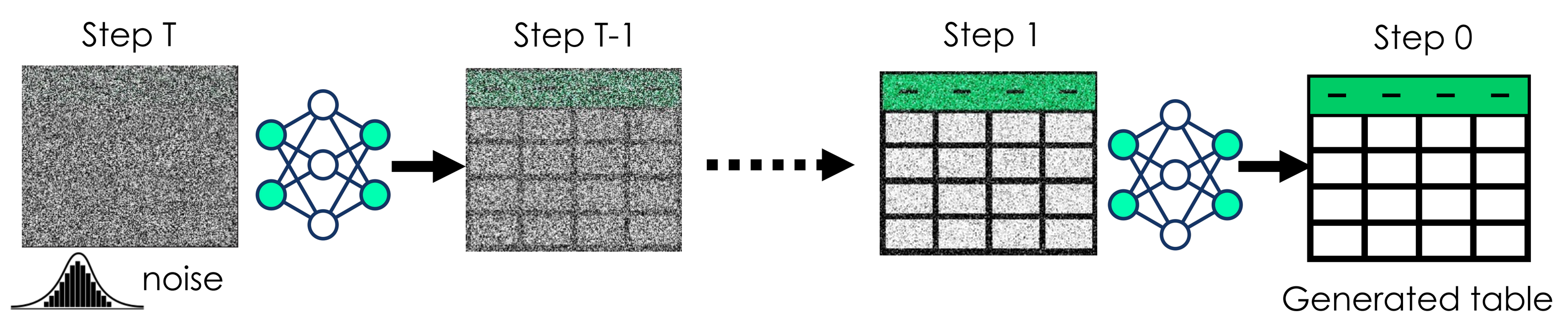
Aditya Shankar [1] Yuandou Wang [1] Rihan Hai [1] Lydia Y. Chen [1, 2]

[1]Delft University of Technology, Delft, The Netherlands [2]University of Neuchâtel, Neuchâtel, Switzerland

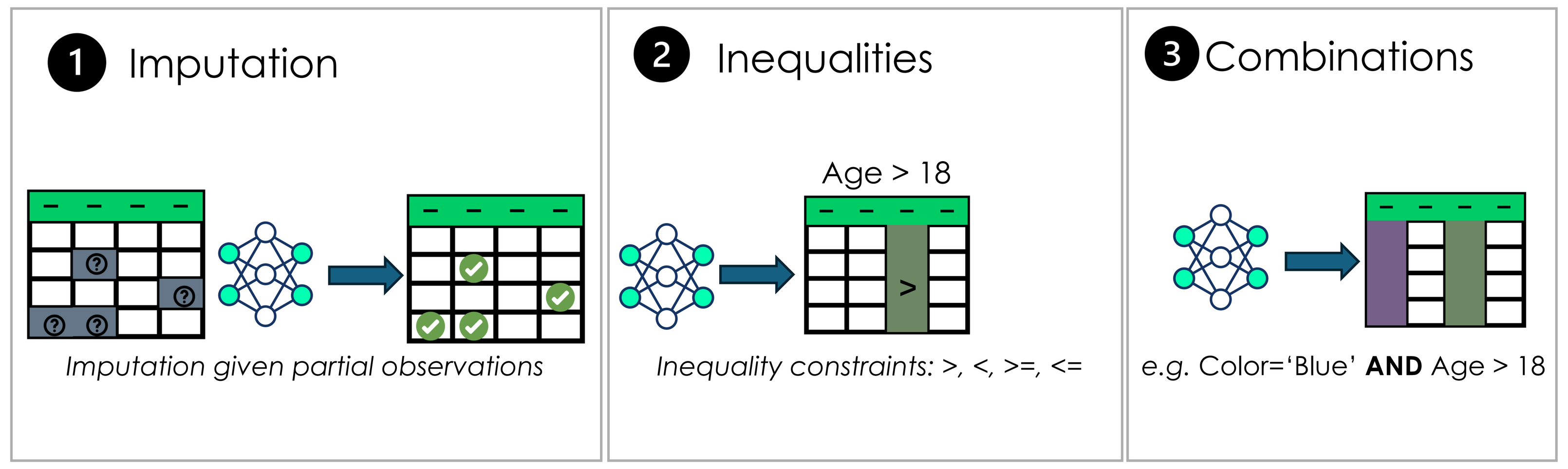
TL;DR: Harpoon generalises manifold theory for handling arbitrary tabular conditional generation tasks at inference time.

Motivation

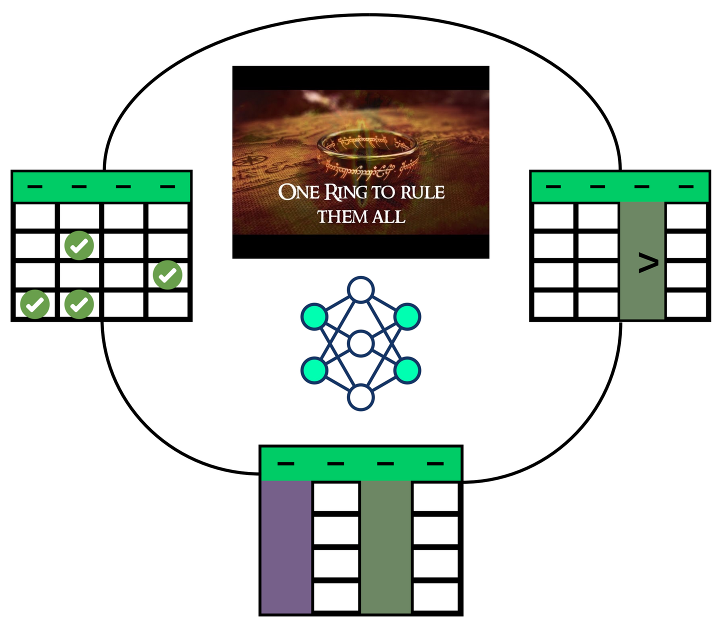
Tabular Diffusion iteratively generates tables from noise.



Conditional tabular diffusion is all about controlling the generative process for various tasks, such as imputation, generating data subject to inequality constraints, or even multiple constraints in combination.

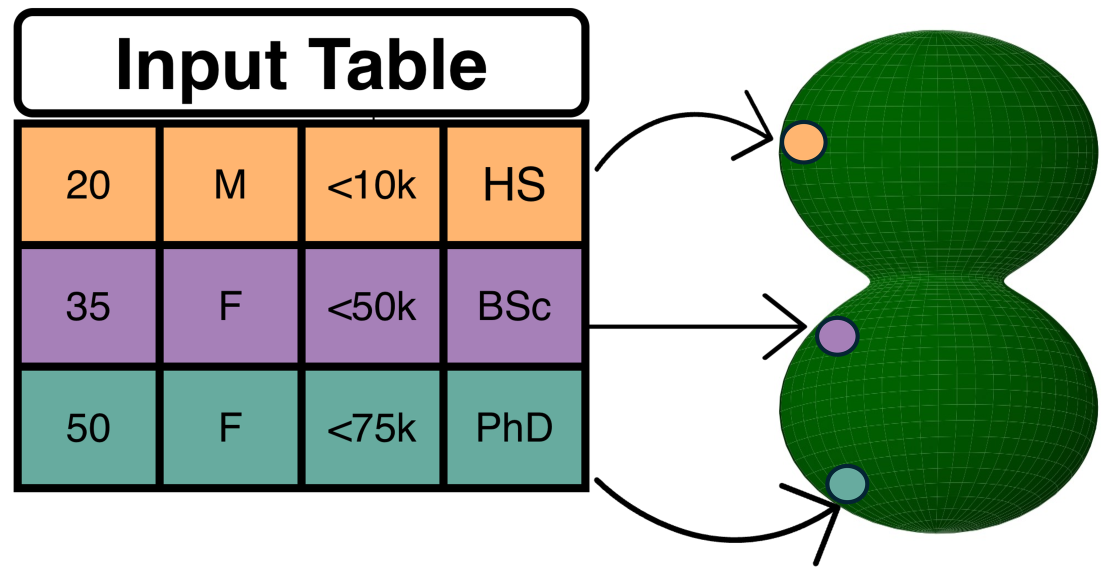


Challenges

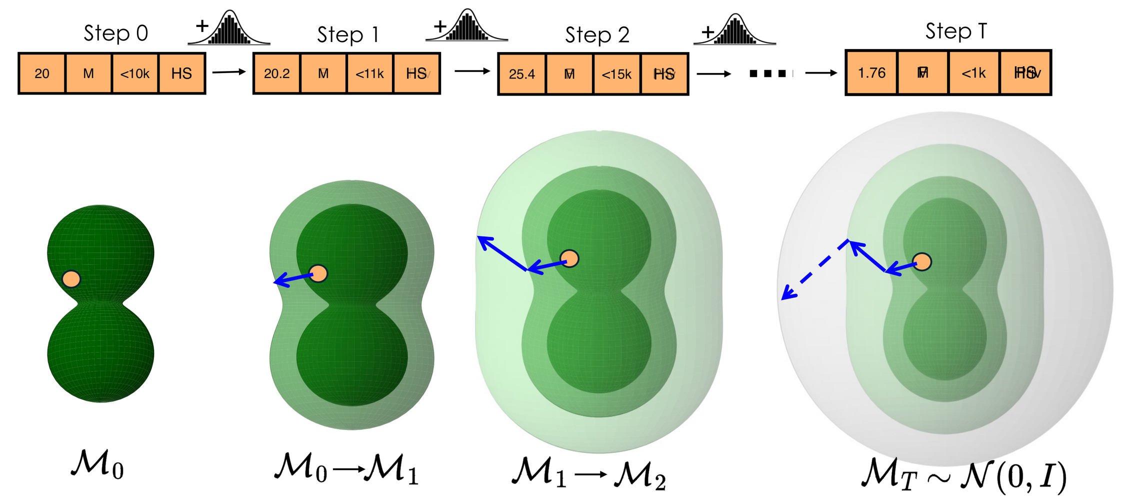


- **One model for all tasks**: We would like to adapt a pre-trained model at **inference** time for various tasks.
- **Theory-to-practice**: We want an approach that is theoretically grounded.
- **Simple and efficient**: We want an approach that is simple to understand and to implement in practice.

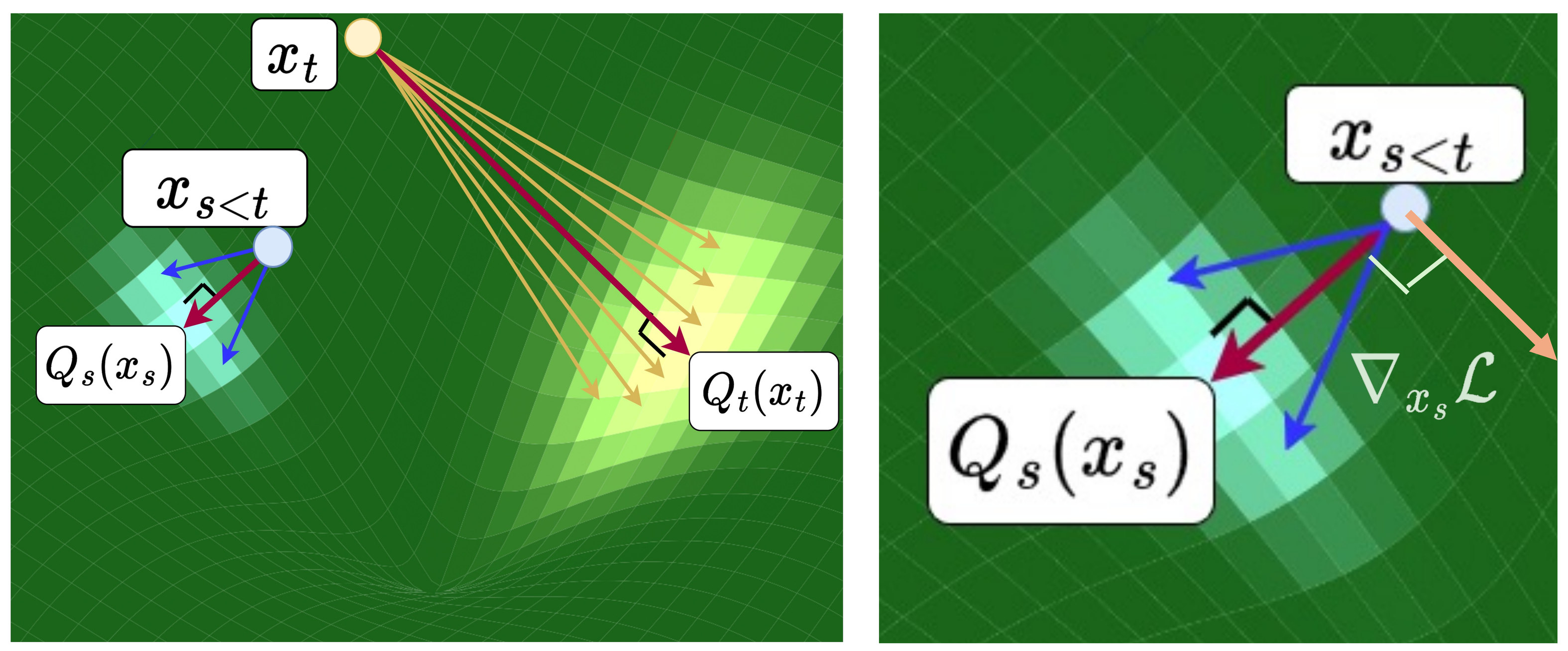
Harpoon: Theory



We assume tabular rows implicitly lie on low-dimensional **manifolds**. Geometrically, diffusion is like expanding the manifold shell until it eventually becomes white noise.

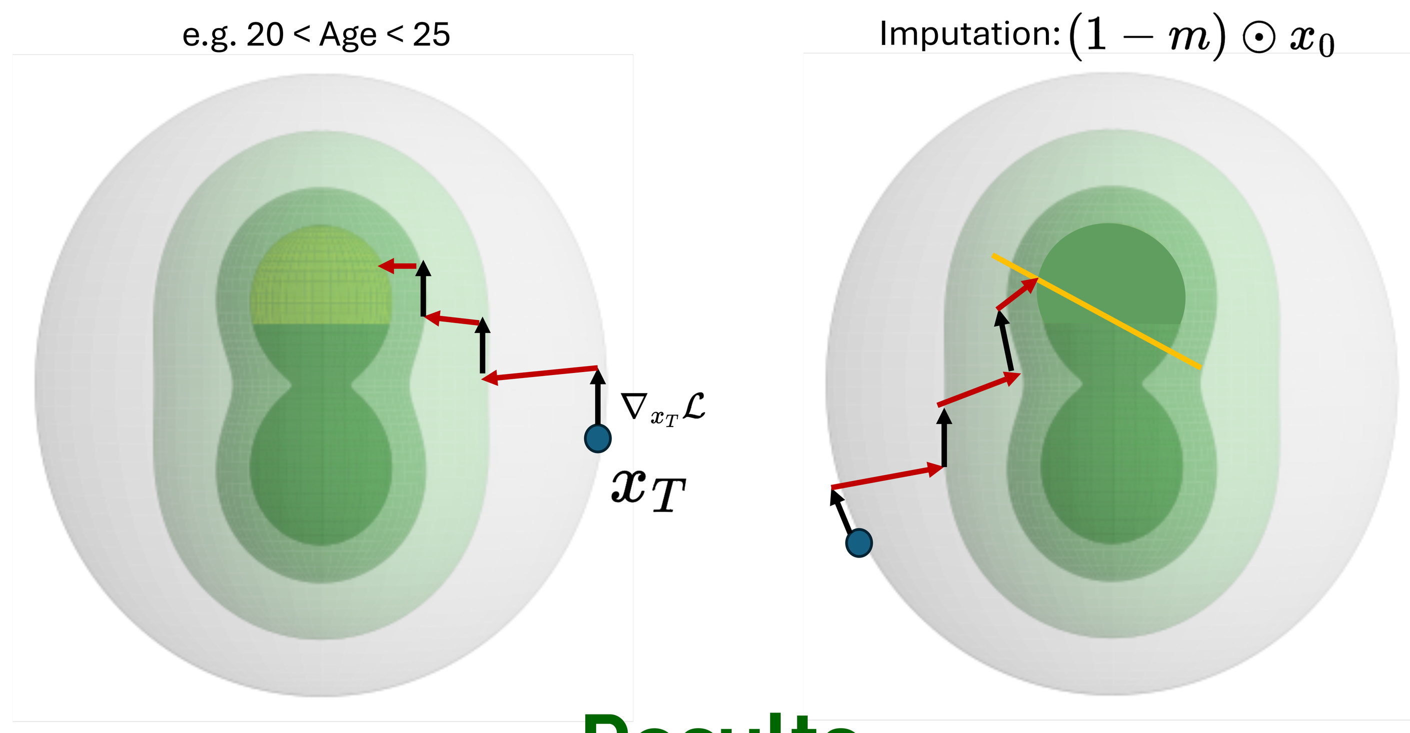


We prove that at small noise levels, diffusion models project **orthogonally** onto the manifold surface. At the same time, gradients from **any** inference time loss align **tangentially** to the surface.



Harpoon: Algorithm

We compute a task-dependent inference-time loss that measures the fidelity with the provided constraints. Over several denoising steps, we **interleave** unconditional denoising steps with tangential corrections to converge.



Results

Harpoon achieves low MSE (continuous features) and high accuracy (discrete features) for imputation tasks. It also achieves low constraint violation rates while generating realistic samples when handling multiple constraints.