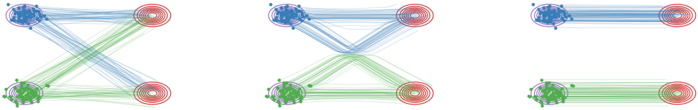
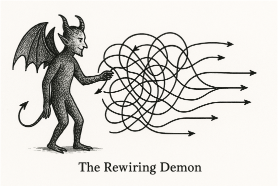
Rewiring Trajectories

- Interpolation paths can intersect and cross
- But trajectories of ODEs can never cross each other.
- Rectified Flow rewires the crossings of interpolation.





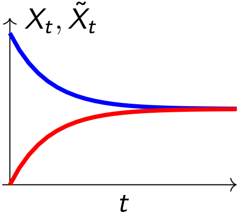
ODEs Trajectories Can Not Cross Each Other

$$\dot{X}_t = v_t(X_t).$$

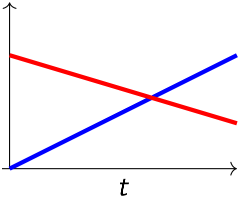
- The update direction $\dot{X}_t$ is uniquely determined by X_t.

Let $\{X_t\}$ and $\{\tilde{X}_t\}$ be solutions of the same ODE. Then

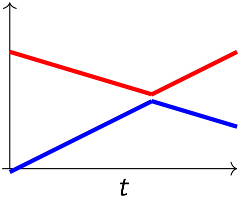
$$X_0 = \tilde{X}_0 \implies X_t = \tilde{X}_t \quad \text{for all } t \text{ in the existence interval.}$$



Possible for ODE



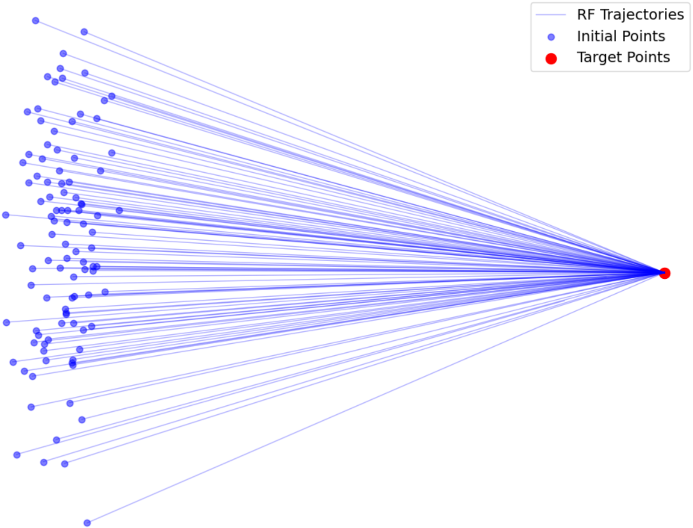
Impossible for ODE



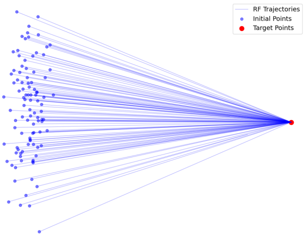
Rewired at crossing

Rectified Flow: Single Data Case

- Consider the case of a single point $x^{\mathtt{data}}$:



Rectified Flow: Single Data Case



- Interpolation:

$$X_t = tx^{\mathtt{data}} + (1-t)X_0.$$

- This interpolation also defines an ODE:

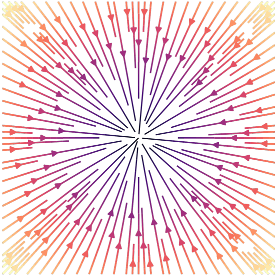
$$\frac{\mathrm{d}}{\mathrm{d}t}X_t = x^{\mathtt{data}} - X_0 = \frac{x^{\mathtt{data}} - X_t}{1-t}.$$

where X_0 is eliminated using the interpolation formula.

$v^*(x,t) = \dfrac{x^{\mathtt{data}} - x}{1-t}$ is the RF velocity field.

Single Point Rectified Flow

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t = \frac{x^{\mathtt{data}} - X_t}{1-t}, \quad t \in [0,1]$$

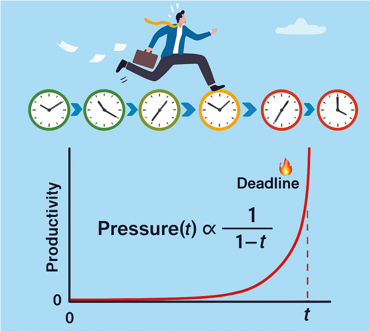


- Apparent singularity from the $1/(1-t)$ factor.
- Yet the solution is perfectly regular and stable:
  - Straight trajectories
  - Finite uniform speed
  - Always arrives at $X_t = x^{\mathtt{data}}$ when $t = 1$
- Also perfectly numerically stable: Euler's method yields exact solution in one step.

Single Point Rectified Flow

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = \frac{x^{\mathtt{data}} - X_t}{1 - t}, \qquad t \in [0, 1]$$

- Intuitively, $1/(1-t)$ is a "deadline pressure".
- Carefully calculated to land $x^{\mathtt{data}}$ precisely at $t = 1$.



Time-Scaled Gradient Flow

- Reparameterize time:

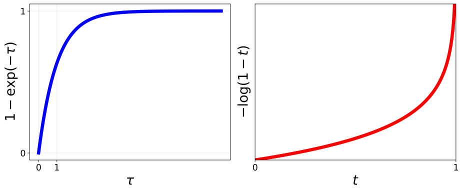
$$\tau = -\log(1-t) \qquad \iff \qquad t = 1 - e^{-\tau}.$$

- Define new variable: $Y_\tau := X_{t(\tau)}$
- Then, the dynamics become:

$$\dot{Y}_\tau = x^{\mathtt{data}} - Y_\tau$$

- This is the standard gradient flow of the quadratic potential:

$$f(y) = \frac{1}{2}\left\| x^{\mathtt{data}} - y \right\|^2$$



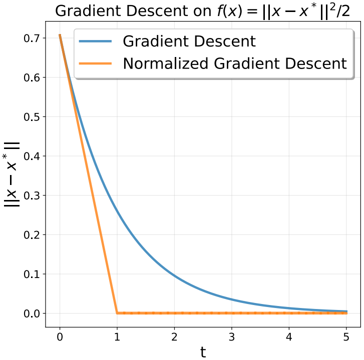
Normalized Gradient Flow

The straight-line ODE $\dot{X}_t = \frac{x^* - X_t}{1-t}$ is also equivalent to

$$\dot{X}_t = -\eta \frac{\nabla f(x)}{\|\nabla f(x)\|},$$

with

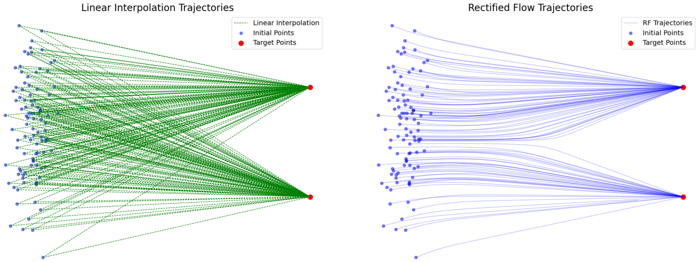
$$f(x) = \frac{1}{2}\|x - x^*\|^2, \quad \eta = \|x_0\|.$$



In general, normalized gradient flow on strongly convex functions [RB20]:

- Normalize the update norm across updates.
- Squeeze gradient flow into finite time.

Rectified Flow: More Data Points



Interpolation Paths

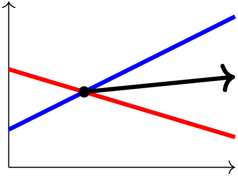
- The interpolated paths have crossings, hence “non-causal”

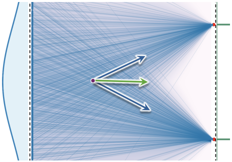
Rectified Flow

- Learns a causal ODE that best approximates the interpolation path.
- Unentangles the path into a forward generative process.
- It de-randomizes, causalizes, and Markovizes the interpolation.

From Interpolation to Generation

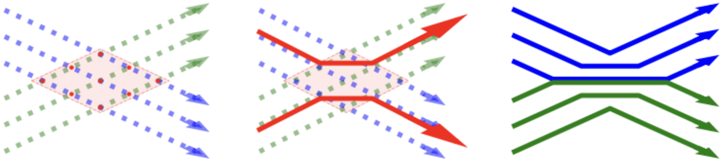
- Projecting the Interpolation Process to the ODE :
$$\min_{v} \mathbb{E}_{(X_0, X_1, t)}\left[\|\dot{X}_t - v_t(X_t)\|^2\right].$$
- The Explicit solution is
$$v^*(x,t) = \mathbb{E}\left[\dot{X}_t \mid X_t = x\right].$$
- The "mean field" velocity: Take the average direction whenever intersection happens.

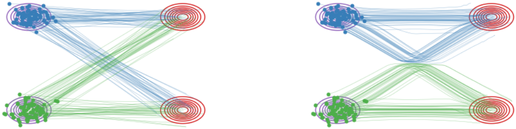




How Does Rewiring Actually Happen by Velocity Averaging?

- How Does Averaging Velocity Lead to Trajectory Rewiring?





Bias-variance Decomposition:

$$
\begin{aligned}
L(v) &= \mathbb{E}\left[\|\dot{X}_t - v_t(X_t)\|^2\right] \\
&= \underbrace{\mathbb{E}\left[\|\dot{X}_t - \mathbb{E}[\dot{X}_t \mid X_t]\|^2\right]}_{\substack{\text{Conditional variance} \\ = \mathbb{E}[\mathrm{Var}(\dot{X}_t|X_t)]}} + \underbrace{\mathbb{E}\left[\|v_t(X_t) - \mathbb{E}[\dot{X}_t \mid X_t]\|^2\right]}_{\text{Estimation bias}}
\end{aligned}
$$

- Hence, the optimal solution should achieve zero bias:

$$v_t^*(X_t) = \mathbb{E}\left[\dot{X}_t \mid X_t\right].$$

Bias-variance Decomposition:

$$L(v) = \mathbb{E}\left[\|\dot{X}_t - v_t(X_t)\|^2\right]$$
$$= \underbrace{\mathbb{E}\left[\|\dot{X}_t - \mathbb{E}[\dot{X}_t \mid X_t]\|^2\right]}_{\substack{\text{Conditional variance} \\ = \mathbb{E}[\mathrm{Var}(\dot{X}_t | X_t)]}} + \underbrace{\mathbb{E}\left[\|v_t(X_t) - \mathbb{E}[\dot{X}_t \mid X_t]\|^2\right]}_{\text{Estimation bias}}$$

- Hence, the optimal solution should achieve zero bias:

$$v_t^*(X_t) = \mathbb{E}\left[\dot{X}_t \mid X_t\right].$$

- The minimum loss value is

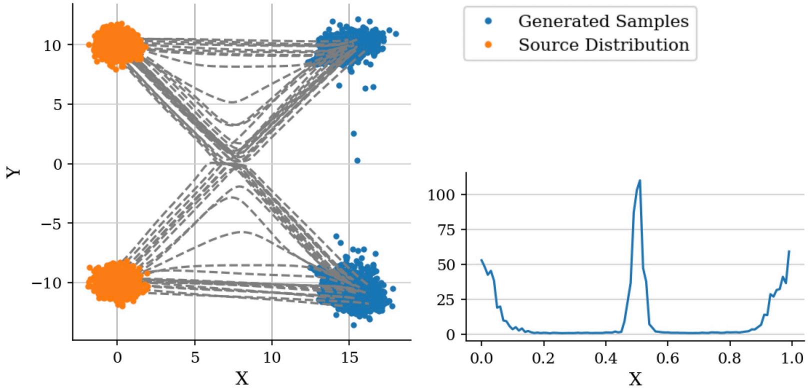
$$L(v^*) = \mathbb{E}\left[\mathrm{Var}(\dot{X}_t \mid X_t)\right].$$

It reflects:

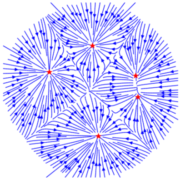
- The **degree of intersection** of interpolation process $\{X_t\}$.
- The **trajectory straightness** of the rectified flow $\{Z_t\}$.

Loss as Straightness

The lower the loss, the **straighter** the ODE path from noise to data.



Singular Velocity on Finite Data Points



On a finite number of data points $\{x^{(i)}\}_{i=1}^{n}$:

$$v^*(x,t) = \sum_{i=1}^{n} \omega_t^{(i)}(x) \left(\frac{x^{(i)} - x}{1 - t} \right),$$

with posterior weights $\omega_t^{(i)}(x) = \frac{\rho_0\left(\hat{x}_0^{(i)} \mid x^{(i)}\right)}{\sum_j \rho_0\left(\hat{x}_0^{(j)} \mid x^{(j)}\right)}$, $\hat{x}_0^{(i)} = \frac{x - tx^{(i)}}{1-t}$.

Finite Mixture of $\frac{x^{(i)}-x}{1-t}$

- Singular velocity due to $1/(1-t)$.
- Dynamics exactly achieves the training data.
- Minimum training loss, but large evaluation loss.
- Neural network **must** provide smoothing as it **can not fit** the $1/(1-t)$ singularity.

Analytic Velocity on Smooth Densities

With smooth densities, we get

$$v_t^*(x) = \mathbb{E}_{X_1 \sim \pi_1}\left[\omega_t(X_1 \mid x)\frac{X_1 - x}{1-t}\right],$$

where $\omega_t(x_1 \mid x)$ is the posterior probability:

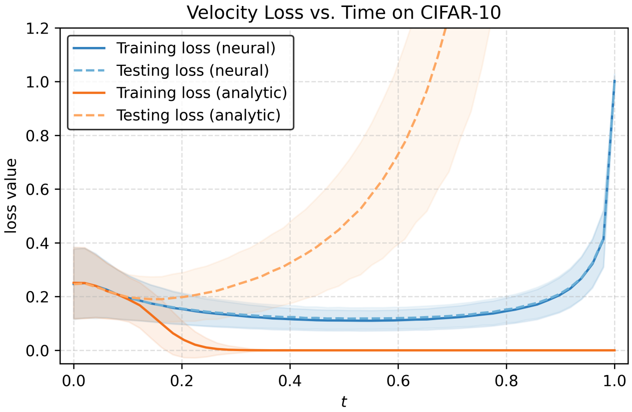
$$\omega_t(x_1 \mid x) := \mathbb{P}(X_1 = x_1 \mid X_t = x) = \frac{\rho_0\left(\hat{x}_0 \mid x_1\right)}{\mathbb{E}_{X_1}\left[\rho_0\left(\hat{X}_0 \mid X_1\right)\right]}, \quad \hat{x}_0 := \frac{x - tx_1}{1-t}$$

where $\rho_0(x_0 \mid x_1)$ is the density of X_0 given X_1.

- **Infinite mixture** of the one-point velocity $\dfrac{x^{\texttt{data}} - x}{1-t}$.
- Singularity may be smoothed out.

Bless of Neural Fitting Error

- The singular analytic velocity on training data fails to generalize.
- But the neural net training refuses the singular solution.
- Avoiding singularity ensures data outside of training set can be sampled, leading to generalization.



Analytic model yields very small training loss yet exploding testing loss.

Open Question:

- Why does neural network generalizes in a way that matches human perception?
- Related: mechanistic explanation of diffusion generalization [NZMW24, SZT17, NBMS17].