Marginal Preservation

$$\boxed{\texttt{Marginals}(\texttt{Rectify}(\{X_t\})) = \texttt{Marginals}(\{X_t\})}$$

The $\texttt{Rectify}(\cdot)$ Operator

Definition. For any stochastic process $\{X_t\}$, its **rectified flow**, denoted as

$$\{Z_t\} = \texttt{Rectify}(\{X_t\}),$$

is the ODE process:

$$\dot{Z}_t = v_t(Z_t), \quad \text{initialized from} \quad Z_0 = X_0,$$

with velocity field

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

Assume the ODE solution exists and unique.

- **Theme**: Understanding and exploiting the $\texttt{Rectify}(\cdot)$ operator.

Basis

- A **random variable** is a measurable map from a random seed ω to a value:

$$X = X(\omega), \qquad \omega \sim \mathsf{P}^{\omega}.$$

- This induces a **distribution** (law) of X, denoted by $\mathsf{P} = \mathrm{Law}(X)$.

- A **stochastic process** is a map from $(\omega, \texttt{time})$ to values:

$$X_t = X(\omega, t).$$

- It induces a **path measure** P on the space of trajectories, and a time marginal $\mathsf{P}_t = \mathrm{Law}(X_t)$.

- The process is said to have **time-differentiable trajectories** if the time derivative exists:

$$\dot{X}_t = \partial_t X(\omega, t) \quad \text{exists.}$$

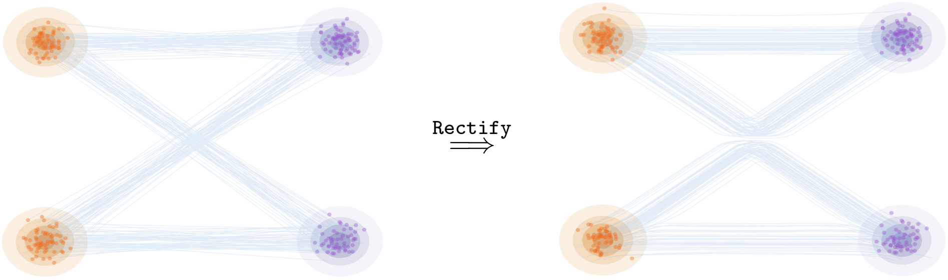
Key Property: Marginal Preservation

Marginal Preservation

The rectified flow preserves marginal distributions:

$$\{Z_t\} = \texttt{Rectify}(\{X_t\}) \quad \Longrightarrow \quad \pi_t = \mathrm{Law}(X_t) = \mathrm{Law}(X_t).$$

Therefore, if (X_0, X_1) is a coupling π, then so is (Z_0, Z_1).



Interpolation: $\{X_t\}$ ODE: $\{Z_t\}$

$$\texttt{Marginals}(\texttt{Rectify}(\{X_t\})) = \texttt{Marginals}(\{X_t\})$$

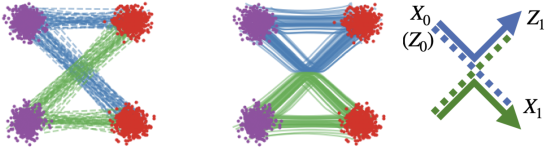
Key Property: Marginal Preservation

$$\boxed{\texttt{Marginals}(\texttt{Rectify}(\{X_t\})) = \texttt{Marginals}(\{X_t\})}$$

Intuition: Marginal Preserving

- Rewiring changes only the flow directions.
- The total flow in/out of each space-time point remains unchanged.

Flow() = Flow(), $\forall$time & location $\Longrightarrow \mathrm{Law}(Z_t) = \mathrm{Law}(X_t), \forall t.$



Continuity Equation: For Smooth Processes

Let $\{X_t\}$ be any stochastic process with smooth trajectories. Define

RF velocity field: $$v_t(x) = \mathbb{E}\left[\dot{X}_t \mid X_t = x\right].$$

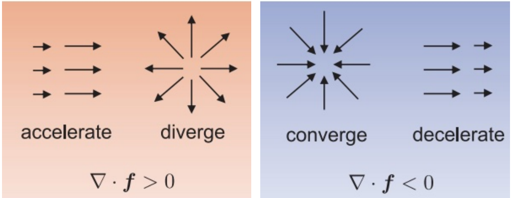
Then the marginal density ρ_t of X_t satisfies the **continuity equation**:

$$\underbrace{\partial_t \rho_t(x)}_{\text{rate of change}} = -\underbrace{\nabla \cdot (v_t(x)\rho_t(x))}_{\text{divergence of flux}}.$$

- Holds for general processes with smooth trajectories: ODE/non-ODE, **Markov/non-Markov**, deterministic/stochastic.

- Divergence of a vector field $v\colon \mathbb{R}^d \to \mathbb{R}^d$:

$$\nabla \cdot v(x) = \mathrm{Trace}(\nabla v(x)).$$



Proof.

For any compactly support, smooth test function h:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[h(X_t)\right] &= \mathbb{E}\left[\nabla h(X_t)^\top \dot{X}_t\right] \\
&= \mathbb{E}\left[\nabla h(X_t)^\top \mathbb{E}\left[\dot{X}_t | X_t\right]\right] && // \quad \text{The law of total expectation } \mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y|X]\right]. \\
&= \mathbb{E}\left[\nabla h(X_t)^\top v_t(X_t)\right] \\
&= \int \nabla h(x)^\top v_t(x)\rho_t(x)\mathrm{d}x \\
&= -\int h(x)\nabla \cdot (v_t(x)\rho_t(x))\mathrm{d}x && // \quad \text{Integration by parts: } \int \nabla h^\top f \mathrm{d}x = -\int h \nabla \cdot f \mathrm{d}x.
\end{aligned}$$

Taking $h = \delta_x$ yields

$$\partial_t \rho_t(x) = -\nabla \cdot (v_t(x)\rho_t(x)).$$

□

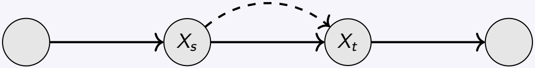
Marginal is a Markovian Property

- Marginals are determined by the **Markov transition probability:**

$$P(X_t \mid X_s) \quad t \geq s.$$

- **Marginals are determined recursively by:**

$$P(X_t) = \int P(X_t \mid X_s) P(X_s) \, \mathrm{d}X_s$$



- Taking the limit $s \to t$ yields density evolution equations.

- It **does not** require the **full transition probability:**

$$\mathrm{P}(X_s \mid X_{\leq t}), \qquad s > t,$$

even if the process itself is non-Markov.

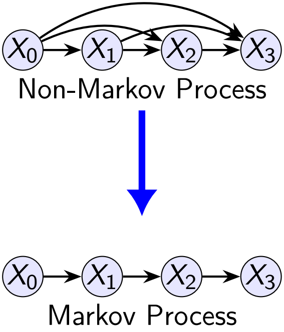
Markovian Projection

Given a joint distribution

$$P^*(X_0, \ldots, X_T) = \prod_t P^*(X_t \mid X_{\mathtt{<t}}),$$

its Markovian projection, or **Markovization** is the solution of

$$\min_P \mathrm{KL}(P^* \,||\, P) \;\; s.t. \;\; P \in \texttt{Markov Chain}.$$



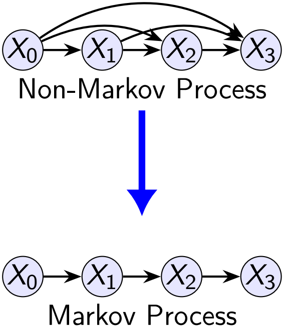
Markovian Projection

Given a joint distribution

$$P^*(X_0, \ldots, X_T) = \prod_t P^*(X_t \mid X_{<\mathtt{t}}),$$

its Markovian projection, or **Markovization** is the solution of

$$\min_P \mathrm{KL}(P^* \mid\mid P) \;\; s.t. \;\; P \in \mathtt{Markov\ Chain}.$$



- **Solution yields**

$$P^{\mathtt{Markov}}(X_0, \ldots, X_T) = \prod_t P^*(X_t \mid X_{\mathtt{t-1}}).$$

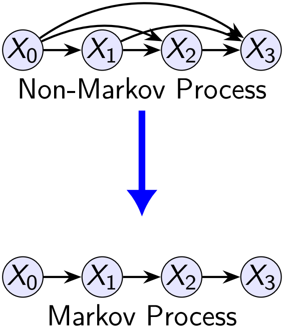
Markovian Projection

Given a joint distribution

$$P^*(X_0, \ldots, X_T) = \prod_t P^*(X_t \mid X_{\color{red}{<t}}),$$

its Markovian projection, or **Markovization** is the solution of

$$\min_P \mathrm{KL}(P^* \,||\, P) \;\; s.t. \;\; P \in \texttt{Markov Chain}.$$



- **Solution yields**

$$P^{\texttt{Markov}}(X_0, \ldots, X_T) = \prod_t P^*(X_t \mid X_{\color{red}{t-1}}).$$

The Markovization preserves the marginals:

$$P^{\texttt{Markov}}(X_t) = P^*(X_t), \quad \forall t.$$

Markovization

The Markovian projection:

$$\min_P \mathrm{KL}(P^* \parallel P) \quad s.t. \quad P \in \texttt{Markov Chain}.$$

Solution yields

$$P^{\texttt{Markov}}(X_0, \ldots, X_T) = \prod_t P^*(X_t \mid X_{\mathtt{t-1}}).$$

Proof.

As Markov chain, we have $P(X_0, \ldots, X_T) = \prod_t P(X_t \mid X_{t-1})$. Hence,

$$\mathrm{KL}(P^* \parallel P) = \sum_t \mathrm{KL}(P^*(X_t \mid X_{t-1}) \parallel P(X_t \mid X_{t-1}))]$$

Minimization yields: $P(X_t \mid X_{t-1}) = P^*(X_t \mid X_{t-1})$. □

Rectification as Markovization

Rectified flow $\{Z_t\} = \texttt{Rectify}(\{X_t\})$ can be viewed as the best Markov approximation of $\{X_t\}$.

Wentzell–Freidlin Principle: Consider the stochastic perturbation

$$\mathrm{d}Z_t^\epsilon = v_t(Z_t^\epsilon)\,\mathrm{d}t + \sqrt{\epsilon}\,\mathrm{d}W_t,$$

the probability of a smooth path $\{x_t\}$ satisfies:

$$\epsilon \log \mathbb{P}(\{Z_t^\varepsilon\} \approx \{x_t\}) \asymp -\int_0^1 \|\dot{x}_t - v_t(x_t)\|^2\,\mathrm{d}t.$$

Hence, the RF loss is asymptotically the KL divergence:

$$\epsilon \mathrm{KL}(\mathbb{P}^X \,\|\, \mathbb{P}^Z) \approx \frac{1}{2}\mathbb{E}\left[\int_0^1 \|\dot{X}_t - v_t(X_t)\|^2\,\mathrm{d}t\right] + const.$$