PROBABILISTIC FORECASTING WITH STOCHASTIC INTERPOLANTS AND FÖLLMER PROCESSES

Yifan Chen*, Mark Goldstein*, Mengjian Hua*, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden.

Courant Institute of Mathematical Sciences (NYU)

PROBABILISTIC FORECASTING

Problem. Forecasting future states of a dynamical system given complete or partial information about current states is ubiquitous in science and engineering

Set-up. Discrete time-series $\{x_{k\tau}\}_{k\in\mathbb{Z}}$ with each $x_{k\tau} \in \mathbb{R}^d$ containing, e.g., daily weather measurements or video frames, acquired every lag-time $\tau > 0$. Successive observations come from joint PDF $\rho(x_{k\tau}, x_{(k+1)\tau})$ with $x_{(k+1)\tau} \sim \rho_c(\cdot|x_{k\tau})$

Stochastic Interpolants. Let x_0 and x_1 denote the current and forecasting state. We introduce the *stochastic interpolant*

$$I_s = \alpha_s x_0 + \beta_s x_1 + \sigma_s W_s \tag{1}$$

where $(x_0, x_1) \sim \rho(x_0, x_1)$ and $W = (W_s)_{s\in[0,1]}$ is a Wiener process with $W \perp (x_0, x_1)$. $\alpha, \beta, \sigma \in C^1([0,1])$ satisfy $\alpha_0 = \beta_1 = 1$ and $\alpha_1 = \beta_0 = \sigma_1 = 0$.

GENERATION WITH STOCHASTIC INTERPOLANTS

Let $b_s(x, x_0)$ be the unique minimizer of

$$L_b[\hat{b}_s] = \int_0^1 \mathbb{E}[|\hat{b}_s(I_s, x_0) - R_s|^2] ds, \tag{2}$$

where $\mathbb{E}$ denotes an expectation over $(x_0, x_1) \sim \rho$ and W with $(x_0, x_1) \perp W$,

$$R_s = \dot{\alpha}_s x_0 + \dot{\beta}_s x_1 + \dot{\sigma}_s W_s. \tag{3}$$

Then the solutions to the SDE

$$dX_s = b_s(X_s, x_0) ds + \sigma_s dW_s, \quad X_{s=0} = x_0, \tag{4}$$

satisfy $\text{Law}(X_s) = \text{Law}(I_s|x_0)$, $\forall (s, x_0) \in [0,1] \times \mathbb{R}^d$. Thus $X_{s=1} \sim \rho_c(\cdot|x_0)$.

KEY OBSERVARIONS

1. Choose the base distribution x_0 as the conditioning information
2. SDE maps a point $X_{s=0} = x_0$ to a density $X_{s=1} \sim \rho_c(\cdot|x_0)$; ODEs cannot
3. Loss function is simulation-free since $W_s \overset{d}{=} \sqrt{s}z$ with $z \sim \mathrm{N}(0, Id)$

GENERALIZATIONS WITH TUNABLE DIFFUSION

Given any $g \in C^0([0,1])$ with mild assumptions, define

$$b_s^g(x, x_0) = b_s(x, x_0) + \tfrac{1}{2}(g_s^2 - \sigma_s^2)\nabla \log \rho_s(x|x_0) \tag{5}$$

where $\rho_s(x|x_0)$ is the PDF of $X_s \overset{d}{=} I_s|x_0$. Then, solutions of

$$dX_s^g = b_s^g(X_s^g, x_0) ds + g_s dW_s, \quad X_{s=0}^g = x_0, \tag{6}$$

satisfy $\text{Law}(X_s^g) = \text{Law}(X_s) = \text{Law}(I_s|x_0)$ for all $(s, x_0) \in [0,1] \times \mathbb{R}^d$. And

$$\nabla \log \rho_s(x|x_0) = A_s [\beta_s b_s(x, x_0) - c_s(x, x_0)], \tag{7}$$

where $A_s = [s\sigma_s(\dot{\beta}_s\sigma_s - \beta_s\dot{\sigma}_s)]^{-1}$ and $c_s(x, x_0) = \dot{\beta}_s x + (\beta_s\dot{\alpha}_s - \dot{\beta}_s\alpha_s)x_0$.

FUNDAMENTAL FACT

We can estimate b first and then adjust both the noise amplitude g_s and the drift b^g *a-posteriori* without having to retrain b.

KL OPTIMIZATION AND FÖLLMER PROCESSES

Question. Is there a g that is special?

Consider the KL between the path measure of $X^g = (X_s^g)_{s\in[0,1]}$ (solving SDE (6)), and the path measure of approximation $\hat{X}^g = (\hat{X}_s^g)_{s\in[0,1]}$ obtained through an estimate $\hat{b}$ of b:

$$D_{\text{KL}}(X^g \| \hat{X}^g) = \int_0^1 \frac{|1 + \frac{1}{2}\beta_s A_s(g_s^2 - \sigma_s^2)|^2}{2|g_s|^2} \mathbb{E}^{x_0}[|\hat{b}_s(I_s, x_0) - b_s(I_s, x_0)|^2] ds \tag{8}$$

Claim. Equation 8 is minimized if we set $g_s = g_s^{\text{F}}$ with

$$g_s^{\text{F}} = \left|2s\sigma_s(\beta_s^{-1}\dot{\beta}_s\sigma_s - \dot{\sigma}_s) - \sigma_s^2\right|^{1/2}. \tag{9}$$

THEOREM

If $\beta_s/[\sqrt{s}\sigma_s]$ is non-decreasing, then the process $X^{g^{\text{F}}}$ is a **Föllmer process**.

- **Föllmer processes** solve the *Schrödinger bridge problem* when one endpoint is a point mass, offering an entropy-regularized solution to optimal transport.
- Usually defined by minimizing KL against the Wiener process subject to constraints on the endpoints
- Our result offers a generalization and new interpretation of Föllmer as the minimizer of the KL of the exact forecasting process from the estimated one, which is more tailored to *statistical inference*.

FORECASTING 2D STOCHASTIC NAVIER STOKES EQUATIONS

Setting : $\mathrm{d}\omega + v\cdot\nabla\omega \mathrm{d}t = \nu\Delta\omega \mathrm{d}t - \alpha\omega \mathrm{d}t + \epsilon \mathrm{d}\eta$ on $\mathbb{T}^2$ where ω is the vorticity, v is the velocity, $\mathrm{d}\eta$ is white-in-time random forcing on a few Fourier modes, and $\nu = 10^{-3}, \alpha = 0.1, \epsilon = 1$. Goal is to forecast $\omega_{t+\tau}$, observing ω_t

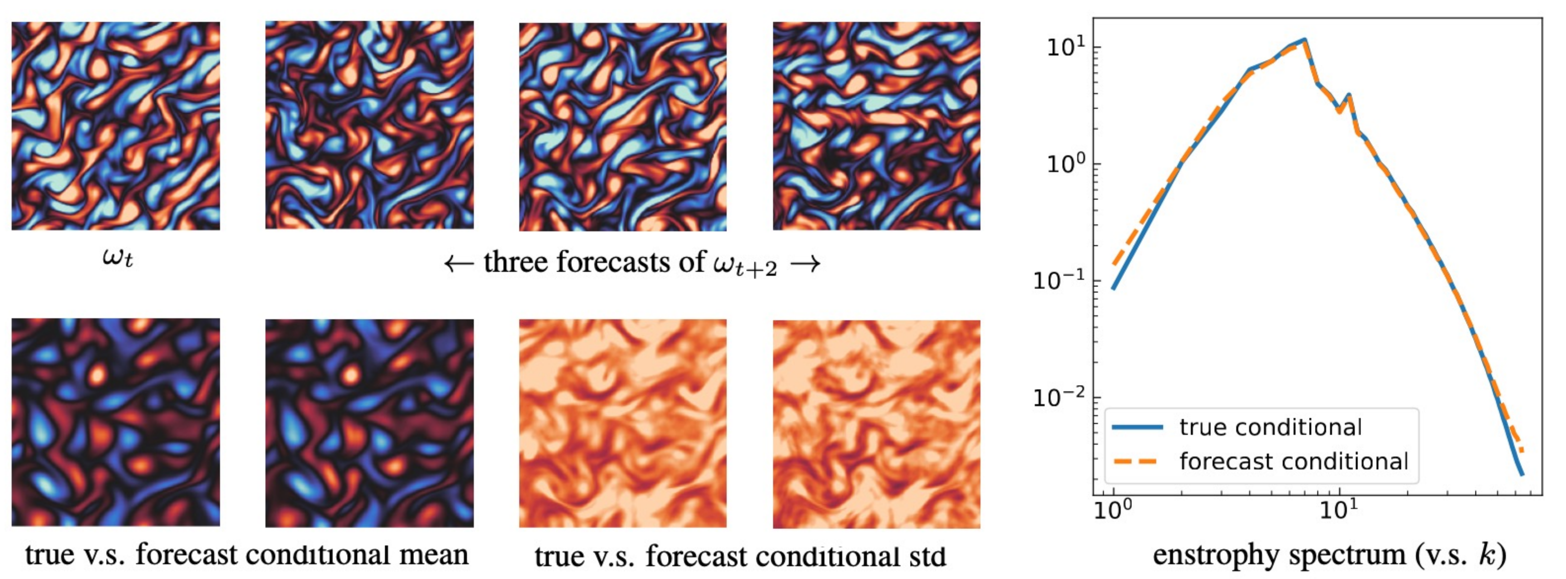

Figure 1. Probabilistic forecasting with lag $\tau = 2$ (autocorrelation 10%). Resolution 128×128, using $200K$ data pairs for training 2M-parameter-Unet for 50 epochs

The effect of tuning g. We examine the total enstrophy and energy statistics

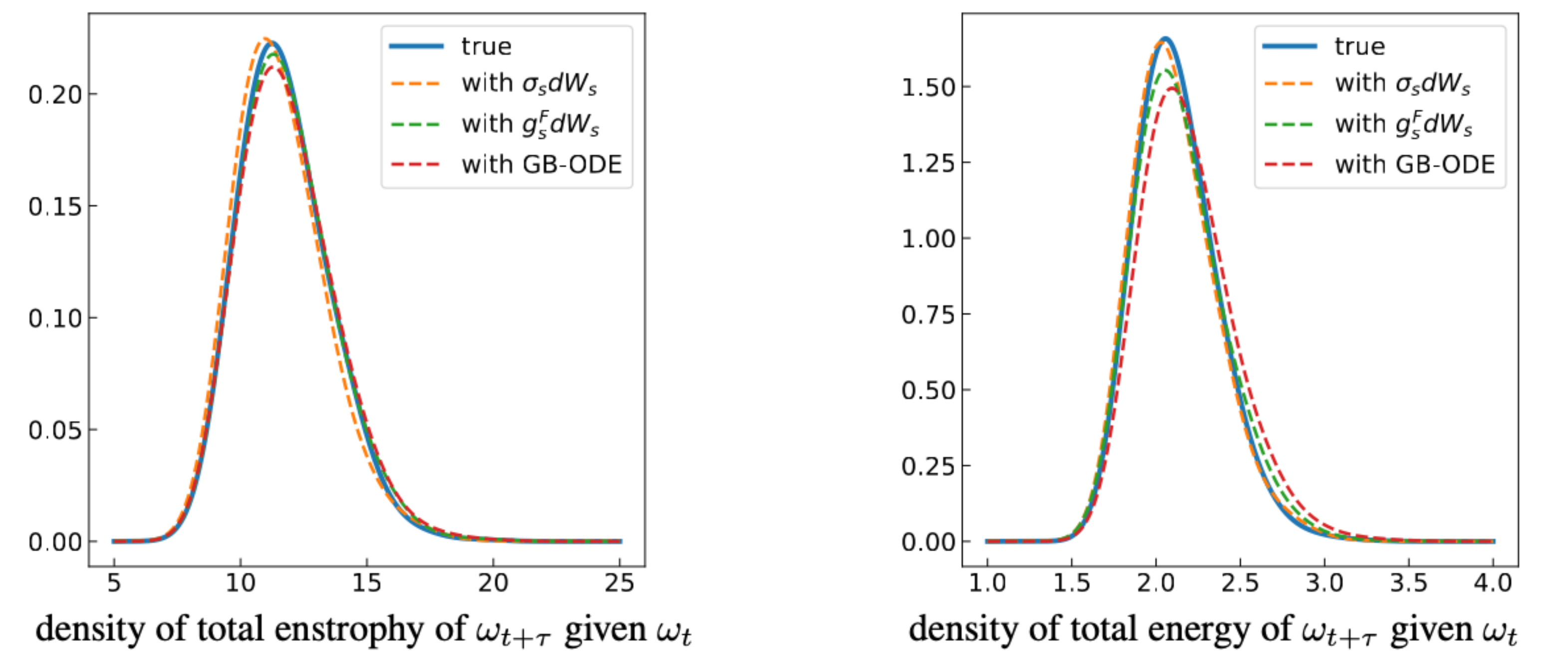

Figure 2. The 1D conditional distributions of total enstrophy and total energy of $\omega_{t+\tau}$, given a fixed initial vorticity field ω_t and $\tau = 1$. Here we compare between the truth, generated samples via SDEs with $\sigma_s dW_s$, via SDE with $g_s^F dW_s$ which corresponds to a Föllmer process, and via ODEs with Gaussian bases

VIDEO FORECASTING

Setting. Video generation on CLEVRER, which features moving geometric objects and their collisions. Following previous work, RIVER, the 128x128 videos are mapped to 16x16 with a VQGAN. We model the latent space with the interpolant. During generation, we condition on 2 real frames and generate 14.

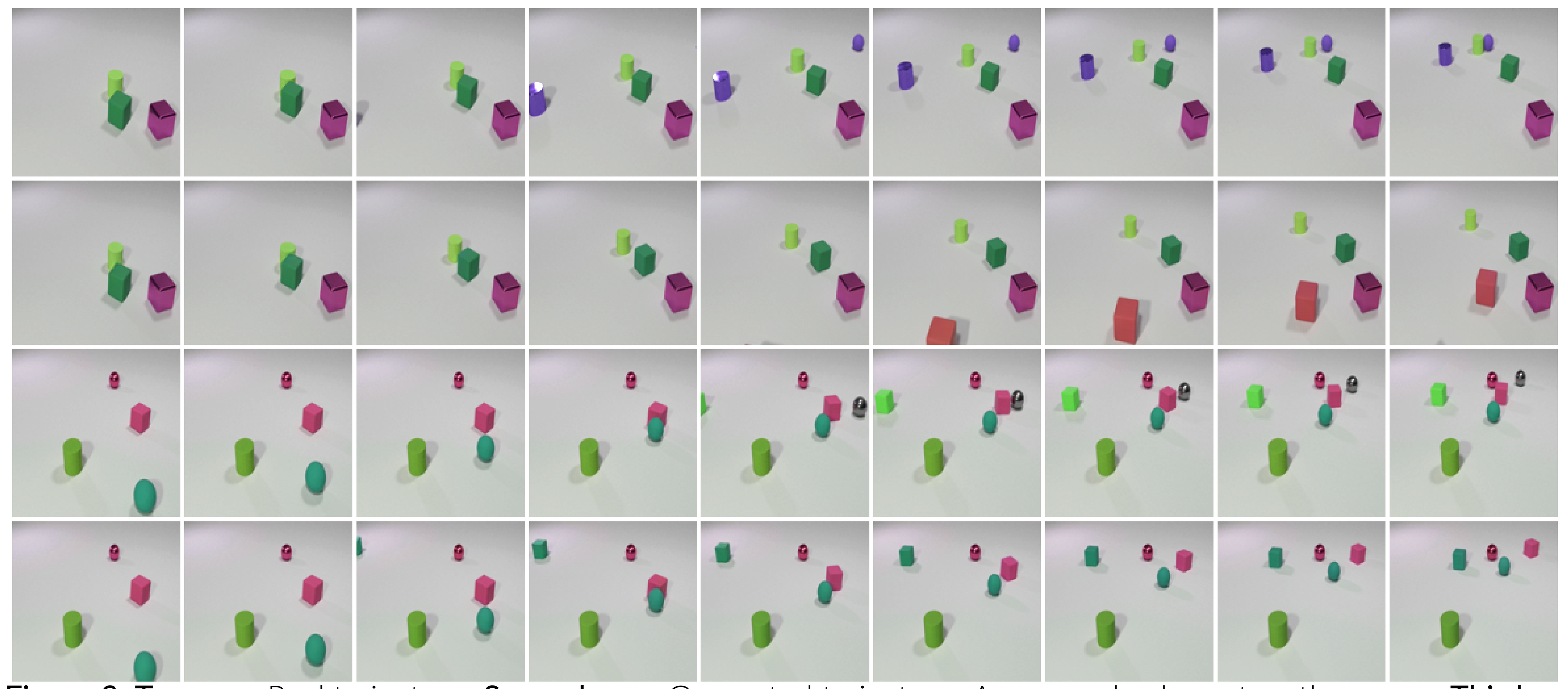

Figure 3. **Top row:** Real trajectory. **Second row:** Generated trajectory. A new, red cube enters the scene. **Third row:** Real trajectory. **Fourth row:** Generated trajectory. A new green cube enters the scene, and collision physics is respected (green ball hits red cube).

Results. The model forecasts video completions, generating new objects and respecting physics (last row, model generates a new green cube in the top left of the scene, which collides realistically with existing objects).

	KTH		*CLEVRER*	
Method	100k	250k	100k	250k
RIVER	46.69	41.88	60.40	48.96
PFI (ours)	**44.38**	**39.13**	**54.7**	**39.31**
Auto-enc.	33.45	33.45	2.79	2.79

Table 1. FVD computed on 256 test set videos, with the model generating 100 completions for each video. Results are reported for 100k grad steps and 250k. The auto-enc represents the FVD of the pretrained encoder-decoder vs the real data. It serves as a bound on the possible model performance, as the modeling is done in the latent space of a pre-trained VQGAN.