FRACTIONAL BROWNIAN BRIDGES FOR ALIGNED DATA

Gabriel Nobis * Fraunhofer HHI
Arina Belova Fraunhofer HHI
Maximilian Springenberg Fraunhofer HHI
Rembert Daems Ghent University FlandersMake–MIRO
Christoph Knochenhauer TU Munich
Manfred Opper TU Berlin University of Potsdam University of Birmingham
Tolga Birdal Imperial College London
Wojciech Samek Fraunhofer HHI TU Berlin BIFOLD

ABSTRACT

Modeling stochastic processes with fractional diffusion instead of purely Brownian-driven dynamics may better account for real-world memory effects, long-range dependencies, and anomalous diffusion phenomena that standard Brownian motion fails to capture. We incorporate fractional Brownian motion (fBM) into aligned diffusion bridges for conformational changes in proteins, utilizing a Markov approximation of fractional Brownian motion (MA-fBM) to study the effect of this generalized prior reference process on predicting future states of the protein conformations from aligned data. We observe that our generalized dynamics yield a lower root mean-squared deviation (RMSD) of C_α atomic positions in the predicted future state from the ground truth. The best performance for this task is achieved with a scaled Ornstein-Uhlenbeck (OU) reference process, which predicts 32% of examples with an RMSD < 2Å on the D3PM test split, whereas a scaled purely Brownian reference process achieves 0% for this threshold.

1 INTRODUCTION

Many real-world systems evolve according to continuous-time, stochastic dynamics [3; 10; 15]. In practice, we only have observations at sparse, discrete time points, e.g. one initial state and one future state. The objective of aligned diffusion Schrödinger bridges (SBALIGN) is to interpolate between these two observations through stochastic trajectories, and to infer future states based on the initial observation. This requires a stochastic model that accurately captures the underlying transition dynamics. The driving noise of SBALIGN is a Brownian motion, which cannot capture correlated increments, memory effects, or anomalous diffusion. In this work, we expand the paired data perspective of Schrödinger bridges proposed in [13] to a driving Markov approximation of fractional Brownian motion (MA-fBM), leading to an approximation of a scaled fractional Brownian bridge for paired data interpolation. We apply our generalized framework to infer conformational changes in proteins and achieve a reduced root mean-squared deviation (RMSD) of C_α atomic positions in the predicted state.

2 BACKGROUND

Our work extends the framework of Somnath et al. [13] to a prior reference process approximating a scaled Riemann-Liouville fractional Brownian motion (fBM) [9]. Assume we observe *i.i.d.* samples of *aligned data* $(\mathbf{x}_0^i, \mathbf{x}_1^i)_{i=1}^N$ drawn from a joint distribution π with marginals $\mathbf{x}_0^i \sim \hat{\mathbb{P}}_0$ and $\mathbf{x}_1^i \sim \hat{\mathbb{P}}_1$. The task is to reconstruct a stochastic process with marginal distribution $(\mathbb{P}_t)_{t\in[0,1]}$ interpolating between $\mathbf{x}_0^i$ and $\mathbf{x}_1^i$ such that $\pi = \mathbb{P}_{0,1}$, where $\mathbb{P}_{0,1}$ is the joint distribution of $\mathbb{P}_0$ and $\mathbb{P}_1$. A key assumption in Somnath et al. [13] is that, given a prior reference process $X = (X_t)_{t\in[0,1]}$ with

*Corresponding author gabriel.nobis@hhi.fraunhofer.de

marginals $(\mathbb{Q}_t)_{t\in[0,1]}$, the minimizer of the Schrödinger bridge problem

$$\pi^\star := \underset{\mathbb{P}_0=\hat{\mathbb{P}}_0,\mathbb{P}_1=\hat{\mathbb{P}}_1}{\arg\min} D_{KL}(\mathbb{P}_{0,1}, \mathbb{Q}_{0,1}) \tag{1}$$

preserves the distribution of the observable data. Here, D_{KL} denotes the Kullback-Leibler divergence, the minimization is taken over all couplings of $\mathbb{P}_0$ and $\mathbb{P}_1$, and $\mathbb{Q}_{0,1}$ denotes the joint distribution of $\mathbb{Q}_0$ and $\mathbb{Q}_1$. This is a strong assumption, as it implies that real-world dynamics follow the dynamics of the minimizing stochastic process *and* that the reference process is rich enough to capture real-world dynamics. While we adopt the first assumption, our work addresses the second by considering a richer class of reference processes that approximate a scaled Riemann-Liouville fractional Brownian motion (fBM) [9], replacing the scaled Brownian motion used in Somnath et al. [13]. See Appendix J for more details on aligned diffusion Schrödinger bridges.

3 Stochastic bridges driven by fractional noise

For ease of notation we assume a data dimension of $d = 1$. The presented theory generalizes to the multidimensional setting, as every data dimension is driven by independent noise. To approximate a reference process driven by Riemann-Liouville (Type II) fractional Brownian motion B_H with Hurst index $H \in (0, 13)$, we replace the Brownian reference process in Somnath et al. [13] by

$$\mathrm{d}X(t) = g(t)\mathrm{d}\hat{B}_H(t), \tag{2}$$

i.e., driven by a Markovian approximation of fBM (MA-fBM), which is defined by the linear superposition of K Ornstein-Uhlenbeck (OU) processes $Y_k(t)$

$$\hat{B}_H(t) := \sum_{k=1}^{K} \omega_k Y_k(t), \quad Y_k(t) := \int_0^t e^{-\gamma_k(t-s)}\mathrm{d}B(s), \tag{3}$$

with weights $\omega_1, \ldots, \omega_K$ and geometrically spaced speeds of mean reversion $\gamma_1(t), ..., \gamma_K(t)$. The augmenting OU processes $Y_1, ..., Y_K$ are all driven by the same standard Brownian motion B, approximating the time-correlated behavior of fBM. For details on the approximation of fBM, we refer the reader to Daems et al. [4]; Harms & Stefanovits [7] and provide a brief summary in Appendix H. In light of the definition of $\hat{B}_H$ in eq. (3), we find that X follows the reference dynamics

$$\mathrm{d}X(t) = -g(t)\sum_{k=1}^{K}\omega_k\gamma_k Y_k(t)\mathrm{d}t + \sum_{k=1}^{K}\omega_k g(t)\mathrm{d}B(t), \quad \mathrm{d}Y_k(t) = -\gamma_k Y_k(t)\mathrm{d}t + \mathrm{d}B(t). \tag{4}$$

Letting $X^{aug} := (X, Y_1, ..., Y_K)^T \in \mathbb{R}^{K+1}$ denote the augmented reference process, it follows that its dynamics can be written more compactly as

$$\mathrm{d}X^{aug}(t) = F(t)X^{aug}(t)\mathrm{d}t + G(t)\mathrm{d}B(t), \tag{5}$$

for a suitable matrix-valued function F and vector-valued function G defined below in eq. (69), and where all $K + 1$ dimensions are driven by the same Brownian motion B.

Partially pinned process To define a stochastic bridge connecting two given data points x_0 and x_1, observe that we only have to steer the first dimension X of the augmented reference process X^{aug} towards x_1, while the terminal values $Y_1(1), \ldots, Y_K(1)$ of the augmenting OU processes are not required to attain a specific value. Conditioning X to attain the terminal value x_1 leads to the partially pinned process $Z_t = X_t^{aug}|(X(0) = x_0, X(1) = x_1)$ with dynamics

$$\mathrm{d}Z(t) = \left[F(t)Z(t) + G(t)G^T(t)u(t, Z(t))\right]\mathrm{d}t + G(t)\mathrm{d}B(t), \tag{6}$$

$$u(t, Z(t)) = [1, \omega_1\zeta_1(t, 1), ..., \omega_K\zeta_K(t, 1)]^T \frac{x_1 - \mu_{1|t}(Z(t))}{\sigma^2_{1|t}}, \tag{7}$$

with $\zeta_k(t, 1) := \int_t^1 -\gamma_k g(s)e^{-\gamma_k(s-t)}\mathrm{d}s$ and where $\mu_{1|t}(z)$ and $\sigma_{1|t}$ denote the conditional mean and conditional variance of the reference process given by

$$\mu_{1|t}(z) := \mathbb{E}\left[X(1)|Z(t) = z\right] = x + \sum_k \omega_k y_k \zeta_k(t, 1), \qquad \sigma^2_{1|t} := Cov(X(1), X(1)|Z_t = z).$$

The above dynamics can be derived either by an optimal stochastic control argument [5] or via an application of Doob's h-transform; see Appendix A for details. As in SBALIGN the bridge process between the two probability measures is constructed by a suitable data average of pinned processes, connecting two individual data points. Hence, the major change is to derive the pinned process for the MA-fBM.

	Wasserstein Distance ↓		
	Moons	**T-Shape**	**Inverse T-Shape**
SBALIGN	0.14 ± 0.01	0.45 ± 0.02	0.37 ± 0.03
FBBALIGN ($H = 0.3$)	0.13 ± 0.02	0.43 ± 0.03	0.38 ± 0.03
OUBALIGN ($H = 0.9$)	$\mathbf{0.12} \pm 0.02$	$\mathbf{0.24} \pm 0.02$	$\mathbf{0.22} \pm 0.02$

Table 1: Mean Wasserstein distance across dimensions averaged over 10 runs.

OUBALIGN	**Wasserstein Distance ↓**								
	$H = 0.9$	$H = 0.8$	$H = 0.7$	$H = 0.6$	$H = 0.5$	$H = 0.4$	$H = 0.3$	$H = 0.2$	$H = 0.1$
Moons	$\mathbf{0.12} \pm 0.02$	0.13 ± 0.03	0.14 ± 0.01	0.16 ± 0.03	0.15 ± 0.02	0.17 ± 0.03	0.15 ± 0.02	0.15 ± 0.02	0.16 ± 0.03
T-Shape	$\mathbf{0.24} \pm 0.02$	0.27 ± 0.01	0.32 ± 0.01	0.37 ± 0.02	0.43 ± 0.02	0.48 ± 0.03	0.53 ± 0.05	0.57 ± 0.05	0.64 ± 0.08
Inverse T-Shape	$\mathbf{0.22} \pm 0.02$	0.26 ± 0.02	0.30 ± 0.02	0.35 ± 0.01	0.40 ± 0.04	0.44 ± 0.02	0.52 ± 0.05	0.50 ± 0.05	0.56 ± 0.07

Table 2: Wasserstein distances using a scaled OU process as a reference process on different toy datasets—unnormalized variance.

Loss function Following Somnath et al. [13], we train two neural networks b_t^θ and m^ϕ by minimizing the regularized loss functional

$$L(\theta, \phi) := \mathbb{E}\left\{ \left\| \int_0^T \frac{x_1 - \mu_{T|t}(Z(t))}{\sigma^2_{T|t}} - \left[b_t^\theta\left(\mu_{T|t}(Z(t)\right) + m_t^\phi\left(\mu_{T|t}(Z(t)\right) \right] \right\|^2 + \lambda_t \| m_t^\phi(\mu_{T|t}(Z(t)) \|^2 \right\}.$$

Note that in contrast to Somnath et al. [13], the input of the two neural networks b_t^θ and m_t^ϕ is taken to be $\mu_{T|t}(Z(t)) = X(t) + \sum_k \omega_k Y_k(t) \zeta_k(t, T)$ instead of $X(t)$.

Ornstein-Uhlenbeck bridge (OUBALIGN) We consider the minimal modification of SBALIGN with a single OU process as a reference process, where we fix the speed of mean reversion at $\gamma_1 = 1$ and $\omega_1 = b/A$ depends on H according to Proposition 2 below. The variance of this reference process is governed by the Hurst index H. Choosing $H > 0.5$ leads to a smaller terminal variance, while $H < 0.5$ results in higher terminal variance.

Fractional Brownian bridge (FBBALIGN) To approximate a Riemann-Liouville fractional Brownian bridge, we use MA-fBM as the driving noise of our stochastic bridge, fixing $K = 5$ throughout all experiments. Following the experimental setup of Somnath et al. [13], we set the diffusion function to a constant $g(t) \equiv g_{max}/\mathbb{V}[\hat{B}_H(T)]$, ensuring a terminal variance of g_{max} for the reference process. In every experiment, we choose g_{max} according to Somnath et al. [13].

4 Experiments

We showcase the performance of our two variants OUBALIGN and FBBALIGN of a generalized Schrödinger bridge for aligned data first on toy data and second to infer conformational changes in proteins. Our implementation is built on the repository provided in Somnath et al. [13].

Toy data In all the toy datasets we have a 8000 to 1000 train to validation split. Our evaluation metric here is the Wasserstein-1 distance [17] (defined in Appendix F) computed between each dimension of the generated and the target test samples, averaged over the two data dimensions. The reported distance is averaged over 10 trials. During one trial, we train two MLPs for 20 epochs and generate 10000 samples via Euler-Maruyama [2] to compute the Wasserstein-1 distance to the target test samples. For realizations of the three considered source (t_0) and target (t_1) distributions "Moons", "T-Shape" and "Inverse T-Shape" from Somnath et al. [13], see Appendix D. The goal on the Moons dataset is to align two half-moon-shaped distributions, each rotated 120 degrees relative to the other. In the task on the T-Shape dataset, one aims to match distributions positioned on different edges of a T-shaped space, while for the Inverse T-Shape dataset, source and target distributions swap roles. For OUBALIGN, we observe in Table 2 that the configuration with the lowest terminal variance under consideration achieves the best performance in terms of average Wasserstein-1 distance across all three

FBBALIGN	**Wasserstein Distance ↓**								
	$H = 0.9$	$H = 0.8$	$H = 0.7$	$H = 0.6$	$H = 0.5$	$H = 0.4$	$H = 0.3$	$H = 0.2$	$H = 0.1$
Moons	0.68 ± 0.46	0.34 ± 0.17	0.30 ± 0.20	0.22 ± 0.05	0.17 ± 0.02	0.14 ± 0.02	$\mathbf{0.13} \pm 0.01$	0.17 ± 0.03	0.27 ± 0.05
T-Shape	1.20 ± 0.97	0.76 ± 0.17	0.62 ± 0.15	0.58 ± 0.13	0.48 ± 0.05	0.47 ± 0.03	$\mathbf{0.43} \pm 0.03$	0.44 ± 0.02	0.63 ± 0.09
Inverse T-Shape	0.69 ± 0.27	0.57 ± 0.11	0.48 ± 0.06	0.49 ± 0.11	0.43 ± 0.04	0.42 ± 0.03	$\mathbf{0.38} \pm 0.03$	0.42 ± 0.03	0.61 ± 0.06

Table 3: Wasserstein distance across different H between generated samples and ground truth on different toy datasts—with normalized terminal variance of $g_{max} = 0.7$ for the Moons dataset and $g_{max} = 1$ for both tasks on the T-Shaped data.

D3PM	H	RMSD(Å)			% RMSD(Å) $< \tau$		
		Median	Mean	Std	$\tau = 2$	$\tau = 5$	$\tau = 10$
SBALIGN	–	3.80	4.98	3.95	0%	69%	93%
OUBALIGN	0.9	**2.98**	**4.30**	3.96	**31**%	**75**%	93%
OUBALIGN	0.8	3.25	4.38	3.98	28%	74%	93%
OUBALIGN	0.7	3.24	4.39	3.78	22%	75%	93%
OUBALIGN	0.6	3.12	4.37	3.77	20%	76%	93%
OUBALIGN	0.5	3.59	4.66	3.84	10%	72%	93%
OUBALIGN	0.4	3.54	4.58	3.66	3%	74%	**95**%
OUBALIGN	0.3	3.69	4.77	3.66	0%	70%	93%
OUBALIGN	0.2	3.71	4.90	3.67	0%	70%	93%

(a) Quantitative results for a single augmenting OU process ($K = 1$; H induces terminal variance).

D3PM	H	RMSD(Å)			% RMSD(Å) $< \tau$		
		Median	Mean	Std	$\tau = 2$	$\tau = 5$	$\tau = 10$
SBALIGN	–	3.80	4.98	3.95	0%	69%	93%
FBALIGN	0.8	3.42	4.66	3.92	2%	**76**%	93%
FBALIGN	0.7	3.59	4.77	3.90	1%	73%	92%
FBALIGN	0.6	3.57	4.80	4.05	2%	74%	92%
FBALIGN	0.5	3.37	4.58	3.76	3%	74%	93%
FBALIGN	0.4	**3.09**	**4.47**	3.73	**4**%	74%	93%
FBALIGN	0.3	3.34	4.50	3.75	3%	74%	93%
FBALIGN	0.2	3.25	4.49	3.68	2%	77%	93%
FBALIGN	0.1	3.47	4.61	3.64	0%	73%	**94**%

(b) Quantitative results for MA-fBM with $K = 5$ augmenting processes.

Table 4: Quantitative results on the D3PM test split.

toy tasks. Varying the Hurst index for FBBALIGN with normalized terminal variance, we observe the best performance for $H = 0.3$ across all three toy datasets—following a U-shaped performance curve over H: monotonically increasing from $H = 0.9$ to $H = 0.3$ and then decreasing from $H = 0.3$ to $H = 0.1$. The performance of the original SBALIGN compared to our methods OUBALIGN and FBBALIGN on the Moons dataset is nearly the same. However, on the T-Shape datasets, OUBALIGN outperforms both FBBALIGN and SBALIGN by a large margin, while FBBALIGN performs better than SBALIGN .

Conformational changes in proteins In this task, we demonstrated the ability of OUBALIGN and FBBALIGN to infer the bound state of a protein from its initial unbound state. We use the D3PM dataset [11], pre-processed and provided in the repository of Somnath et al. [13]. The dataset consists of 1,591 protein pairs, each containing an unbound and a bound state. We followed the same train/validation/test split as the original work, with 1,291, 150 and 150 examples, respectively. The evaluation metrics include the mean, median, and standard deviation of the RMSD (defined in Appendix F) between the carbon atoms of the predicted and ground truth bound protein structures. Following Somnath et al. [13], we report the percentage of generated proteins with an RMSD below $2.0, 5.0, 10.0$ Å compared to the true structures. The best result on the protein conformations task was achieved using OUBALIGN with $H = 0.9$, based on the mean/median RMSD and the percentage of predicted and ground truth molecules aligned within 2 Å. For the FBBALIGN framework, the best results for most statistical measures were obtained with $H = 0.4$.

5 Conclusion

We extend the framework of aligned diffusion Schrödinger bridges to a prior reference process approximating a scaled fBM and present two variations of this generalized Schrödinger bridge for paired data: OUBALIGN with an OU process as a reference process and FBBALIGN with a driving MA-fBM. On toy data and in the task of predicting conformational changes in proteins, we observe the best overall performance is achieved with a OU process as a reference process with the lowest terminal variance under consideration. For FBBALIGN, the terminal variance is identical across all Hurst indices and can be disregarded for the sake of comparison. Since we observe the best performance for FBBALIGN across all tasks at $H = 0.3$ and $H = 0.4$, we conjecture that a path rougher than a Brownian path is preferable. To investigate this conjecture further, a first step is to retrain all configurations using the empirically best-performing terminal variance for all configuration of FBBALIGN .

Limitations While we are confident in our results on toy data, our conclusions for predicting the bound state of proteins are based on a very limited number of experiments, as we do not average our quantitative results over multiple runs. Due to limited computational resources, each configuration was trained only once.

Meaningfulness Statement

A meaningful representation of life should capture the stochastic processes that drive biological dynamics. Our work advances this direction by generalizing the driving noise in aligned diffusion Schrödinger Bridges from a Brownian motion to a Markovian approximation of fractional Brownian motion, providing increased flexibility for modeling biological variability. Importantly, we do not claim that real-world biological dynamics are driven by fractional Brownian motion, but rather that our approach makes a broader range of dynamics accessible to interpolate between snapshots of biological processes.

REFERENCES

[1] Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*. Springer-Verlag London Limited 2008, 2008. doi:10.1007/978-1-84628-797-8.

[2] Samuel N. Cohen and Robert J. Elliott. *Stochastic Calculus and Applications*. Probability and Its Applications. Birkhäuser, New York, NY, 2st edition, 2015. ISBN 978-1-4939-2866-8.

[3] Rama Cont. Long range dependence in financial markets. In Jacques Lévy-Véhel and Evelyne Lutton (eds.), *Fractals in Engineering*, pp. 159–179, London, 2005. Springer London.

[4] Rembert Daems, Manfred Opper, Guillaume Crevecoeur, and Tolga Birdal. Variational inference for SDEs driven by fractional noise. In *The Twelfth International Conference on Learning Representations*, 2024.

[5] Rembert Daems, Manfred Opper, Guillaume Crevecoeur, and Tolga Birdal. Efficient training of neural sdes using stochastic optimal control. In *ESANN*, 2025.

[6] Philipp Harms. Strong convergence rates for Markovian representations of fractional processes. *Discrete and Continuous Dynamical Systems - B*, 26(10):5567–5579, 2021. ISSN 1531-3492.

[7] Philipp Harms and David Stefanovits. Affine representations of fractional processes with applications in mathematical finance. *Stochastic Processes and their Applications*, 129(4):1185–1228, 2019. ISSN 0304-4149.

[8] Dang Huy. A remark on non-Markov property of a fractional Brownian motion. *Vietnam Journal of Mathematics*, 31, 01 2003.

[9] Paul Lévy. Random functions: general theory with special reference to Laplacian random functions. *University of California Publications in Statistics*, 1:331–390, 1953.

[10] Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian Motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

[11] Cheng Peng, Xinben Zhang, Zhijian Xu, Yanqing Yang, Tingting Cai, and Weiliang Zhu. D3pm: a comprehensive database for protein motions ranging from residue to domain. *BMC Bioinformatics*, 23, 02 2022. doi:10.1186/s12859-022-04595-0.

[12] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks, 2022. URL https://arxiv.org/abs/2102.09844.

[13] Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned diffusion schrödinger bridges. In *The 39th Conference on Uncertainty in Artificial Intelligence*, 2023. URL https://openreview.net/forum?id=BkWFJN7_bQ.

[14] Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*, volume 10. Cambridge University Press, 2019.

[15] MURAD S. TAQQU, VADIM TEVEROVSKY, and WALTER WILLINGER. Estimators for long-range dependence: An empirical study. *Fractals*, 03(04):785–798, 1995. doi:10.1142/S0218348X95000692.

[16] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds, 2018. URL https://arxiv.org/abs/1802.08219.

[17] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

A Dynamics of the partially pinned process

Daems et al. [5] use a Gaussian expression for the reference process to construct the posterior SDE that is steered towards x_1. We derive for a fixed data pair (x_0, x_1) the dynamics of the partially pinned process $X^{aug}(t)|(X(0) = x_0, X(1) = x_1)$ using Doob's h-transform, resulting in the same dynamics as found in Daems et al. [5]. Towards that goal, we define the transform

$$h : [0,1] \times \mathbb{R}^{K+1} \to [0,1], \quad (t,z) \mapsto q(x_1|z(t)), \tag{8}$$

where $q(\cdot|z(t))$ satisfies

$$\mathbb{P}(X_1 \in A|Z_t = z(t)) = \int_A q(x|z(t))dx, \quad A \subset \mathbb{R}. \tag{9}$$

Denote by $p(z,t) = p(x, \mathbf{y}, t) = p(x, y^1, ..., y^K, t)$ the density of Z_t such that

$$q(x) = \int_{\mathbb{R}^K} p(x, \mathbf{y}, 1)\mathrm{d}\mathbf{y} \tag{10}$$

and write

$$p(\tilde{z}(t+s)|z(t)) = p(\tilde{z}, t+s|z, t) \tag{11}$$

for the transition density of Z from time t to $t+s$. To show that h defined in eq. (8) satisfies the space-time regularity property we mimic the proof of [14, Theorem 7.11]. We write with Bayes rule

$$p(z(t+s)|z(t), x_1) = \frac{q(x_1|z(t+s), z(t))p(z(t+s)|z(t))}{q(x_1|z(t))} \tag{12}$$

$$= \frac{q(x_1|z(t+s))p(z(t+s)|z(t))}{q(x_1|z(t))}, \tag{13}$$

where we use for the second equation that Z is a Markov process. Hence, equivalently

$$p(z(t+s)|z(t))q(x_1|z(t+s)) = p(z(t+s)|z(t), x_1)q(x_1|z(t)), \tag{14}$$

such that

$$\int_{\mathbb{R}^{K+1}} p(z(t+s)|z(t))h(t+s,z)\mathrm{d}z(t+s) = \int_{\mathbb{R}^{K+1}} p(z(t+s)|z(t))q(x_1|z(t+s))\mathrm{d}z(t+s) \tag{15}$$

$$\overset{(14)}{=} \int_{\mathbb{R}^{K+1}} p(z(t+s)|z(t), x_1)q(x_1|z(t))\mathrm{d}z(t+s) \tag{16}$$

$$= q(x_1|z(t)) \underbrace{\int_{\mathbb{R}^{K+1}} p(z(t+s)|z(t), x_1)\mathrm{d}z(t+s)}_{=1} \tag{17}$$

$$\overset{(8)}{=} h(t,z). \tag{18}$$

Hence, by Särkkä & Solin [14, eq. (7.73) - eq. (7.78)] we conclude that the partially pinned process $Z(t) := X^{aug}(t)|(X(0) = x_0, X(1) = x_1)$ follows the dynamics

$$\mathrm{d}Z(t) = \left[F(t)Z(t) + G(t)G^T(t)\nabla_z \log q(x_1|Z(t))\right]\mathrm{d}t + G(t)\mathrm{d}B_t. \tag{19}$$

B Explicit calculation of the h-transform

To further specify the above defined transform $h(t,z) = q(x_T|z(t))$ we calculate for the reference process $X(T)$ conditioned on $X(t) = x$ and $Y_k(t) = y_k$ at some time t, for future time T (Note that

the derivations are valid for both Type I and II MA-fBM, since it is conditioned on the current state):

$$X(T) = x + \int_t^T g(s)\mathrm{d}\hat{B}_H(s) \tag{20}$$

$$= x + \sum_k \omega_k \int_t^T g(s)\left(-\gamma_k Y_k(s)\mathrm{d}s + \mathrm{d}B(s)\right) \tag{21}$$

$$= x + \sum_k \omega_k \int_t^T g(s)\left(-\gamma_k \left(y_k e^{-\gamma_k(s-t)} + \int_t^s e^{-\gamma_k(s-r)}\mathrm{d}B(r)\right)\mathrm{d}s + \mathrm{d}B(s)\right) \tag{22}$$

$$= x + \sum_k \omega_k \left(\int_t^T -\gamma_k y_k g(s) e^{-\gamma_k(s-t)}\mathrm{d}s + \int_t^T \int_t^s -\gamma_k g(s) e^{-\gamma_k(s-r)}\mathrm{d}B(r)\mathrm{d}s + \int_t^T g(s)\mathrm{d}B(s)\right) \tag{23}$$

$$= x + \sum_k \omega_k \left(\int_t^T -\gamma_k y_k g(s) e^{-\gamma_k(s-t)}\mathrm{d}s + \int_t^T \int_r^T -\gamma_k g(s) e^{-\gamma_k(s-r)}\mathrm{d}s\mathrm{d}B(r) + \int_t^T g(s)\mathrm{d}B(s)\right) \tag{24}$$

$$= x + \sum_k \omega_k \left(\int_t^T -\gamma_k y_k g(s) e^{-\gamma_k(s-t)}\mathrm{d}s + \int_t^T \int_s^T -\gamma_k g(r) e^{-\gamma_k(r-s)}\mathrm{d}r\mathrm{d}B(s) + \int_t^T g(s)\mathrm{d}B(s)\right) \tag{25}$$

$$= x + \sum_k \omega_k \left(\int_t^T -\gamma_k y_k g(s) e^{-\gamma_k(s-t)}\mathrm{d}s + \int_t^T \int_s^T -\gamma_k g(r) e^{-\gamma_k(r-s)}\mathrm{d}r + g(s)\mathrm{d}B(s)\right) \tag{26}$$

$$= x + \sum_k \omega_k \left(y_k \zeta_k(t,T) + \int_t^T \zeta_k(s,T) + g(s)\mathrm{d}B(s)\right), \tag{27}$$

where for convenience $\zeta_k(t,T) = \int_t^T -\gamma_k g(s) e^{-\gamma_k(s-t)}\mathrm{d}s$. This leads for $z = (x, y_1, ..., y_K)$ to

$$\mu_{T|t}(z) := \mathbb{E}\left[X(T)|X(t) = x, Y_k(t) = y_k\right] = x + \sum_k \omega_k y_k \zeta_k(t,T) \tag{28}$$

and

$$\sigma^2_{T|t} = \mathrm{Cov}\left(X(T_1), X(T_2)|Z(t) = z\right) \tag{29}$$

$$= \sum_{k,l} \omega_k \omega_l \int_t^{\min(T_1,T_2)} \left(\zeta_k(s,T_1) + g(s)\right)\left(\zeta_l(s,T_2) + g(s)\right)\mathrm{d}s \tag{30}$$

does not depend on the condition $z = (x, y_1, ..., y_K)$, where $t < T_1$ and $t < T_2$. Moreover, since

$$Y_k(T) = Y_k(0)e^{-T\gamma_k} + \int_0^T e^{-\gamma_k(T-s)} dB_s \tag{31}$$

$$= e^{-\gamma_k(T-t)}\left[Y_k(0)e^{-t\gamma_k} + \int_0^t e^{-\gamma_k(t-s)} dB_s\right] + \int_t^T e^{-\gamma_k(T-s)} dB_s \tag{32}$$

$$= e^{-\gamma_k(T-t)} Y_k(t) + \int_t^T e^{-\gamma_k(T-s)} dB_s, \tag{33}$$

we have

$$\mu^k_{T|t}(z) := \mathbb{E}\left[Y(T)|X(t) = x, Y_k(t) = y_k\right] = e^{-\gamma_k(T-t)} y_k(t) \tag{34}$$

and

$$\mathrm{Cov}\left(X(T_1), Y_l(T_2)|Z(t) = z\right) = \sum_{k=1}^K \omega_k \int_t^{\min\{T_1,T_2\}} (\zeta_k(s,T_1) + g(s)) e^{-\gamma_l(T_2-s)} ds. \tag{35}$$

does not depend on the condition z as well. Hence, we calculate

$$\partial_x \log q(x_T|z) = -\frac{x_T - \mu_{T|t}(z)}{\sigma^2_{T|t}} \left(-\partial_x \mu_{T|t}(z)\right) = \frac{x_T - \mu_{T|t}(z)}{\sigma^2_{T|t}} \tag{36}$$

and

$$\partial_{y^k} \log q(x_T|z) = -\frac{x_T - \mu_{T|t}(z)}{\sigma^2_{T|t}} \left(-\partial_{y^k} \mu_{T|t}(z)\right) = \left[\omega_k \zeta_k(t,T)\right] \frac{x_T - \mu_{T|t}(z)}{\sigma^2_{T|t}}, \tag{37}$$

such that the explicit dynamics of the partially pinned process are given by

$$\mathrm{d}Z_t = \left[F(t)Z_t + G(t)G^T(t)u(t,Z_t)\right]\mathrm{d}t + G(t)\mathrm{d}B_t, \tag{38}$$

$$u(t,Z_t) = \left[1, \omega_1\zeta_1(t,T), ..., \omega_K\zeta_K(t,T)\right]^T \frac{x_T - \mu_{T|t}(Z(t))}{\sigma^2_{T|t}}. \tag{39}$$

C SAMPLING FROM PARTIALLY PINNED PROCESS

Partially pinned process For $s < t < T$ we have

$$(X(t), Y_1(t), ..., Y_K(t), X(T)|Z(s) = z)^T \sim \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\mu}_{t|s}(z)\\ \mu_{T|s}(z)\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{t|s} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \sigma^2_{T|s}\end{pmatrix}\right), \tag{40}$$

with

$$\boldsymbol{\mu}_{t|s}(z) = (\mu_{t|s}(z,s), \mu^1_{T|t}(z), ..., \mu^K_{t|s}(z))^T \tag{41}$$

and

$$\boldsymbol{\Sigma}_{12} = (cov(X(t), X(T)), cov(Y_1(t), X(T)), ..., cov(Y_K(t), X(T)))^T = \boldsymbol{\Sigma}^T_{21}. \tag{42}$$

Hence, the process partially pinned at (x_s, x_T) follows the distribution

$$Z(t)|(X(s) = x_s, X(T) = x_T) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}), \tag{43}$$

with

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}_{t|s}(z) + \frac{1}{\sigma^2_{T|s}}\boldsymbol{\Sigma}_{12}(x_T - \mu_{T|s}(z)) \tag{44}$$

$$\overset{s=0}{=} \begin{cases} (x_0, 0, ..., 0)^T + \frac{1}{\sigma^2_{T|0}}\boldsymbol{\Sigma}_{12}(x_T - x_0 - \sum_k \omega_k y_0 \zeta_k(0,T)), & \text{(Type I)}\\ (x_0, 0, ..., 0)^T + \frac{1}{\sigma^2_{T|0}}\boldsymbol{\Sigma}_{12}(x_T - x_0), & \text{(Type II)}\end{cases} \tag{45}$$

and

$$\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \frac{1}{\sigma^2_{T|t}}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma} - \frac{1}{\sigma^2_{T|t}}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^T_{12}. \tag{46}$$

We further calculate for a constant diffusion coefficient $g(t) \equiv g \in \mathbb{R}$

$$\zeta_k(s,t) = \int_s^t -\gamma_k g(u) e^{-\gamma_k(u-s)}\mathrm{d}u = -g\gamma_k \int_s^t e^{-\gamma_k(u-s)}du = g(e^{-\gamma_k(t-s)} - 1) \tag{47}$$

and

$$\mu_{T|t}(z) = x + \sum_k \omega_k y_k \zeta_k(t,T) = x + g\sum_{k=1}^K \omega_k(e^{-\gamma_k(T-t)} - 1)y_k, \tag{48}$$

$$\mu^k_{T|t}(z) = e^{-\gamma_k(T-s)}y_k. \tag{49}$$

Left to calculate are the entries of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{1,2}$. With $s < t \leq T$ we calculate

$$\mathrm{Cov}(X(t), X(T)|Z(s) = z) = \sum_{i,j=1}^K \omega_i\omega_j \int_s^t \left(\zeta_i(u,t) + g\right)\left(\zeta_j(u,T) + g\right)\mathrm{d}u \tag{50}$$

$$= g^2 \sum_{i,j=1}^K \omega_i\omega_j \int_s^t \left((e^{-\gamma_k(t-u)} - 1) + 1\right)\left((e^{-\gamma_k(T-u)} - 1) + 1\right)\mathrm{d}u \tag{51}$$

$$= g^2 \sum_{i,j=1}^K \omega_i\omega_j \int_s^t \left(e^{-\gamma_i(t-u)}\right)\left(e^{-\gamma_j(T-u)}\right)\mathrm{d}u \tag{52}$$

$$\text{Cov}(Y_i(t), Y_j(T)|Z(s) = z) = \int_s^t e^{-\gamma_i(t-u)} e^{-\gamma_j(T-u)} du \tag{53}$$

$$= \frac{e^{-T\gamma_j - t\gamma_i} \left(e^{t\gamma_j + t\gamma_i} - e^{s\gamma_j + s\gamma_i}\right)}{\gamma_j + \gamma_i}, \tag{54}$$

$$\text{Cov}(Y_l(t), X(T)) = \sum_{k=1}^{K} \omega_k \int_0^t e^{-\gamma_l(t-u)} (\zeta_k(u, T) + g(u)) du \tag{55}$$

$$= g \sum_{k=1}^{K} \omega_k \int_0^t e^{-\gamma_l(t-u)} e^{-\gamma_k(T-u)} du \tag{56}$$

$$= g \sum_{k=1}^{K} \omega_k \frac{\left(\mathrm{e}^{t(\gamma_l+\gamma_k)} - 1\right) \mathrm{e}^{-t\gamma_l - T\gamma_k}}{\gamma_l + \gamma_k}. \tag{57}$$

D TOY DATASETS

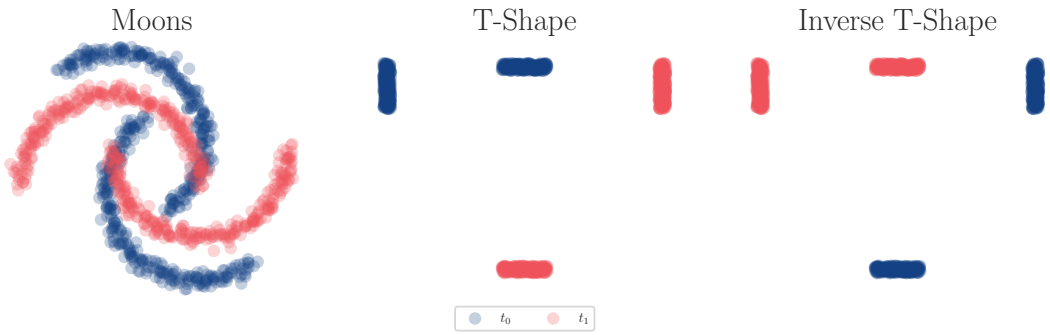


Figure 1: Toy datasets from Somnath et al. [13]. Realizations of source (red) and target (blue) distribution.

E TRAINING AND EVALUATION OF CONFORMATIONAL CHANGES IN PROTEINS DATASET

To evaluate the models on this dataset we simulated the process for 100 steps. Moreover, the training and evaluation dataset was augmented by generating 10 SE(3) symmetrical variants of every protein to confirm that our models respect these symmetries.

F EVALUATION METRICS

Wasserstein distance To measure the distance from the original data distribution from the predicted data distribution we use Wasserstein-1 distance. The Wasserstein-1 distance between ground truth data distribution p_t and sampled data distribution p_s is defined as

$$W_1(p_t, p_s) = \inf_{\gamma \sim \Pi(p_t, p_s)} \mathbb{E}_{(x,\hat{x})}[||x - \hat{x}||]. \tag{58}$$

The lower the Wasserstein distance, the better are the distributions p_t and p_s aligned.

Root Mean Square Deviation Root mean square deviation of C_α atomic positions is a distance between two superimposed molecules/proteins. If $\mathbf{x}$ is an observed 3D structure/configuration of the protein and $\hat{\mathbf{x}}$ is a predicted configuration of the protein then

$$\text{RMSD}(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} ||x_i - \hat{x}_i||^2}. \tag{59}$$

The lower the RMSD, the lower their L2-distance w.r.t. some unit of measure. In our example, the unit of the measure is Angstrom, Å.

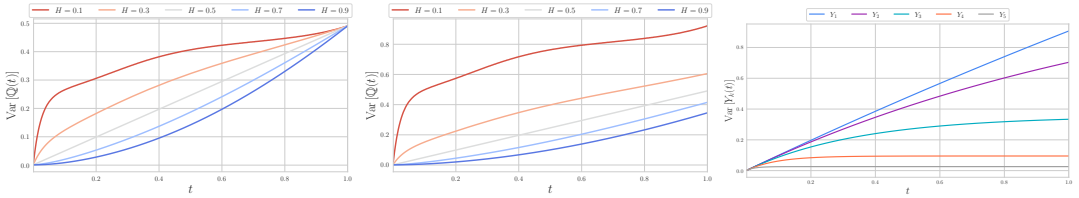
G NETWORK ARCHITECTURES

Baseline network We started off with network architecture proposed by [13]. With introduction of fractional Brownian noise we needed to change the input to the networks.

Toy datasets Here we use two Multi Layer Perceptrons (MLP) to approximate drift $b^\theta(t, \mathbf{Z}_t)$ and Doob's h-score $m^\phi(t, b_t, \mathbf{Z}_t)$.

Conformational changes in proteins Here we used one graph neural network with tensor product convolution layers introduced by [16]. This neral network approximates both the drift and Doob's h-score.

H MARKOVIAN APPROXIMATION OF FRACTIONAL BROWNIAN MOTION



(a) Variance of the reference process X with normalized terminal variance through $g(t) \equiv 1/\mathbb{V}[\hat{B}_H(T)]$.

(b) Variance of the unnormalized reference process X with $g(t) \equiv 1$, recovering MA-fBM.

(c) Variance of the augmenting OU processes $Y_1, ..., Y_5$ approximating fBM as a weighted sum.

Figure 2: Variance evolution of MA-fBM over time.

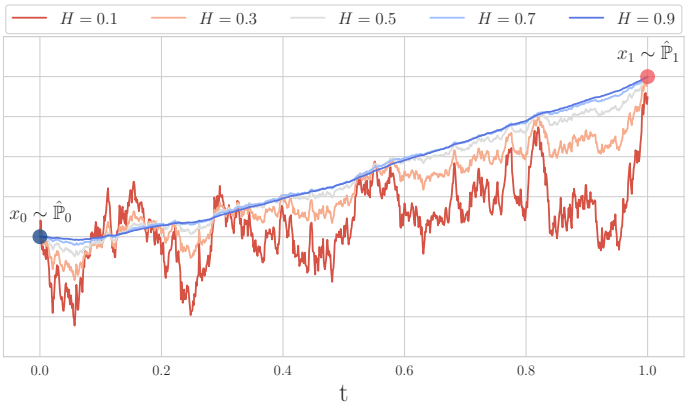


Figure 3: Visualization of the relation between hurst index H and the smoothness of paths with MA-fBM.

The key challenge of fBM is that fBM lacks the Markov property and is not a semimartingale for $H \neq 0.5$ [1; 8]. To address this, we approximate fBM using a linear combination of Markovian semimartingales, leveraging an infinite-dimensional Markovian representation [4; 7]. In Definition 1 we outline the Markovian approximation, as proposed by Daems et al. [4]; Harms & Stefanovits [7], in Proposition 2 we describe how to obtain a solution for MA-fBM following Daems et al. [4].

Definition 1 (Markovian approximation of fBM [4; 6])**.** *Choose $K \in \mathbb{N}$ Ornstein–Uhlenbeck (OU) processes*

$$Y_k(t) = \int_0^t e^{-\gamma_k(t-s)} \mathrm{d}B_s, \quad k \in \mathbb{N}, \quad t \geq 0, \tag{60}$$

with speeds of mean reversion $\gamma_1, ..., \gamma_K$ and dynamics $\mathrm{d}Y_k(t) = -\gamma_k Y_k(t)\mathrm{d}t + \mathrm{d}B_t$. Given a Hurst index $H \in (0, 1)$ and a geometrically spaced grid $\gamma_k = r^{k-n}$ with $r > 1$ and $n = \frac{K+1}{2}$ we call the

process

$$\hat{B}_H(t) := \sum_{k=1}^{K} \omega_k Y_k(t), \quad t \geq 0, \tag{61}$$

Markov-approximate fractional Brownian motion (MA-fBM) with approximation coefficients $\omega_1, ..., \omega_K \in \mathbb{R}$ *and denote by* $\hat{\mathbf{B}}_H = (\hat{B}_H(1), ..., \hat{B}_H(D))^T$ *the corresponding* D*-dimensional process where* $\hat{B}_H(i)$ *and* $\hat{B}_H(j)$ *are independent for* $i \neq j$ *inheriting independence from the underlying standard BMs* B_i *and* B_j*.*

Proposition 2 (Optimal Approximation Coefficients, Daems et al. [4])**.** *The optimal approximation coefficients* $\boldsymbol{\omega} = (\omega_1, ..., \omega_K) \in \mathbb{R}^K$ *for a given Hurst index* $H \in (0, 1)$*, a terminal time* $T > 0$ *and a fixed geometrically spaced grid to minimize the* $L^2(\mathbb{P})$*-error*

$$\mathcal{E}(\boldsymbol{\omega}) := \int_0^T \mathbb{E}\left[\left(B_H(t) - \hat{B}_H(t)\right)^2\right] \mathrm{d}t \tag{62}$$

are given by the closed-form expression $\boldsymbol{A\omega} = \boldsymbol{b}$ *with*

$$\boldsymbol{A}_{i,j} := \frac{2T + \frac{e^{-(\gamma_i+\gamma_j)T}-1}{\gamma_i+\gamma_j}}{\gamma_i + \gamma_j}, \quad \boldsymbol{b}_k := \frac{T}{\gamma_k^{H+\frac{1}{2}}} P\left(H + \frac{1}{2}, \gamma_k T\right) - \frac{H+\frac{1}{2}}{\gamma_k^{H+\frac{3}{2}}} P\left(H + \frac{3}{2}, \gamma_k T\right) \tag{63}$$

and where $P(z, x) = \frac{1}{\Gamma(z)} \int_0^x t^{z-1} e^{-t} \mathrm{d}t$ *is the regularized lower incomplete gamma function.*

Approximating fractional Schrödinger Bridges To approximate a reference process driven by type II fractional Brownian motion $\mathbf{B}^H$ we approximate the reference process by

$$\mathrm{d}X(t) = g(t)\mathrm{d}\hat{\mathbf{B}}_H(t) \approx g(t)\mathrm{d}\mathbf{B}^H(t), \quad t \in [0, 1], \tag{64}$$

where every dimension has a driving MA-fBm $\hat{B}_H(t)$ in accordance with Definition 1

$$\hat{B}_H(t) := \sum_{k=1}^{K} \omega_k Y_k(t) \quad \textit{with } Y_k(t) := \int_0^t e^{-\gamma_k(t-s)} \mathrm{d}B_s. \tag{65}$$

This yields the augmented prior dynamics

$$\mathrm{d}X(t) = g(t) \sum_{k=1}^{K} \omega_k \mathrm{d}Y_k(t) = -g(t) \sum_{k=1}^{K} \omega_k \gamma_k Y_k(t)\mathrm{d}t + \bar{\boldsymbol{\omega}} g(t)\mathrm{d}B_t, \tag{66}$$

$$dY_k(t) = -\gamma_k Y_k(t)\mathrm{d}t + \mathrm{d}B_t, \tag{67}$$

with $\bar{\boldsymbol{\omega}} := \sum_{k=1}^{K} \omega_k$. Resulting, for every data dimension, in the $K + 1$ dimensional augmented reference process $X^{aug} = (X, Y_1, ..., Y_K)$ with dynamics

$$\mathrm{d}X^{aug}(t) = F(t)X^{aug}(t)\mathrm{d}t + G(t)\mathrm{d}B_t, \tag{68}$$

where $F(t) \in \mathbb{R}^{(K+1)\times(K+1)}$ and $G(t) \in \mathbb{R}^{K+1}$ are given by

$$F(t) = \begin{pmatrix} 0 & -g(t)\omega_1\gamma_1 & -g(t)\omega_2\gamma_2 & \dots & -g(t)\omega_K\gamma_K \\ 0 & -\gamma_1 & 0 & \dots & 0 \\ 0 & 0 & -\gamma_2 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & \dots & -\gamma_K \end{pmatrix}, \quad G = \begin{pmatrix} \bar{\boldsymbol{\omega}} g(t) & 1 & \dots & 1 \end{pmatrix}^T. \tag{69}$$

I Related work

Our work builds upon [13], which established a Schrödinger bridge model for aligned data and our implementation is based on their repository.[1] Equivariant graph neural networks (EGNN) [12] is a baseline for protein conformational changes. It is a model that respects roto-translational invariance of the molecules which is important for the task since it removes these unnecessary degrees of freedom.

[1] https://github.com/vsomnath/aligned_diffusion_bridges

J ALIGNED DIFFUSION SCHRÖDINGER BRIDGES

The reference process in [13] is a scaled Brownian motion

$$\mathrm{d}X_t = g(t)\mathrm{d}\mathbf{B}_t, \tag{70}$$

with a driving d-dimensional Brownian motion $\mathbf{B} = (\mathbf{B}_t)_{t\in[0,T]}$ and $g : [0, 1] \to \mathbb{R}^+$. Conditioning the reference process on the aligend data pair $(\mathbf{x}_0, \mathbf{x}_T) \sim \pi^\star$ results in the scaled Brownian bridge $\hat{\mathbf{X}} = \mathbb{Q}|(\mathbf{x}_0, \mathbf{x}_T)$ following the dynamics

$$\mathrm{d}\hat{\mathbf{X}}_t = g_t^2 \frac{\mathbf{x}_T - \hat{\mathbf{X}}_t}{\beta_1 - \beta_t} dt + g_t \mathrm{d}\mathbf{B}_t, \quad \hat{\mathbf{X}}_0 = \mathbf{x}_0, \quad \beta_t = \int_0^t g_s^2 \mathrm{d}s. \tag{71}$$

The scaled Brownian bridge eq. (71) is used in a second step to train the neural networks b^θ and m^ϕ via the loss function

$$L(\theta, \phi) := \mathbb{E}\left[\left\| \int_0^T \frac{\mathbf{x}_T - \hat{\mathbf{X}}_t}{\beta_1 - \beta_t} - (b^\theta(\hat{\mathbf{X}}_t, t) - m^\phi(\hat{\mathbf{X}}_t)) \right\|^2\right] \tag{72}$$

aiming for $b^\theta(\cdot, t) \approx b_t$ and $m^\phi \approx \nabla \log h_t$ to approximate the unknown dynamics

$$\mathrm{d}\mathbf{X}_t^\star = g_t^2 \left[b_t(\mathbf{X}_t^\star) + \nabla \log h_t(\mathbf{X}_t^\star)\right] \mathrm{d}t + g_t \mathrm{d}\mathbf{B}_t, \quad h_t(x) := \mathbb{P}(\mathbf{X}_T = \mathbf{x}_T | \mathbf{X}_t = \mathbf{x}) \tag{73}$$

of the process minimizing

$$\pi^\star := \underset{\mathbb{P}_0=\hat{\mathbb{P}}_0, \mathbb{P}_1=\hat{\mathbb{P}}_1}{\arg\min} D_{KL}(\mathbb{P}_{0,1}, \mathbb{Q}_{0,1}), \tag{74}$$

where we minimize over all couplings of $\mathbf{X}_0 \sim \mathbb{P}_0$ and $\mathbf{X}_T \sim \mathbb{P}_1$ and $\mathbb{Q}_{0,1}$ is the joint distribution of $\mathbb{Q}_0$ and $\mathbb{Q}_1$.