A Theoretical results

Quantity	Description	Introduced in
$\boldsymbol{A} \in \mathbb{R}^{d\times d}, \boldsymbol{m} \in \mathbb{R}^{d}$	Multivariate Ornstein-Uhlenbeck drift parameters	Eq. (1)
$\boldsymbol{D} = \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top$	Multivariate Ornstein-Uhlenbeck diffusivity	Eq. (1)
$\mathbb{P}$	Schrödinger bridge path measure	Section 2.1
$\mathbb{Q}$	Reference process path measure	Section 2.1
$\boldsymbol{\mu}_t^{\boldsymbol{x}_0}, \boldsymbol{\Sigma}_t$	Unconditional mean and covariance of mvOU process started at $\boldsymbol{x}_0$	Section 3.1
$\boldsymbol{p}_{t\mid(\boldsymbol{x}_0,\boldsymbol{x}_T)}$	Conditional density of mvOU bridge	Section 3.1
$\boldsymbol{c}_{t\mid(\boldsymbol{x}_0,\boldsymbol{x}_T)}$	mvOU bridge control between $(\boldsymbol{x}_0, \boldsymbol{x}_T)$	Section 3.1
$\boldsymbol{\mu}_{t\mid(\boldsymbol{x}_0,\boldsymbol{x}_T)}, \boldsymbol{\Sigma}_{t\mid(\boldsymbol{x}_0,\boldsymbol{x}_T)} = \boldsymbol{\Omega}_t$	Conditional mean and covariance of mvOU bridge	Section 3.1
$\boldsymbol{\Omega}_{st}$	Conditional covariance process of mvOU bridge	Section 3.1
$\boldsymbol{u}_{t\mid(\boldsymbol{x}_0,\boldsymbol{x}_T)}, \boldsymbol{s}_{t\mid(\boldsymbol{x}_0,\boldsymbol{x}_T)}$	Conditional probability flow and score field of mvOU bridge	Section 3.1
$\mathcal{N}(\boldsymbol{a}, \boldsymbol{\mathcal{A}}), \mathcal{N}(\boldsymbol{b}, \boldsymbol{\mathcal{B}})$	Initial and terminal mvOU-GSB marginals	Section 3.2
$\mathcal{N}(\overline{\boldsymbol{a}}, \overline{\boldsymbol{\mathcal{A}}}), \mathcal{N}(\overline{\boldsymbol{b}}, \overline{\boldsymbol{\mathcal{B}}})$	Transformed mvOU-GSB marginals	Section 3.2
$\overline{\boldsymbol{\mathcal{C}}}$	Cross-covariance of entropic transport plan	Section 3.2
$\mathfrak{A}_t, \mathfrak{B}_t, \mathfrak{c}_t$	Key quantities for the mvOU-GSB	Section 3.2
$\boldsymbol{\nu}_t$	mvOU-GSB mean process	Section 3.2
$\boldsymbol{\Xi}_t, \boldsymbol{\Xi}_{st}$	mvOU-GSB variance and covariance process	Section 3.2
$\boldsymbol{S}_t^\top \boldsymbol{\Xi}_t^{-1}$	SDE drift matrix of mvOU-GSB	Section 3.2

Table 3: Glossary of some key notations and quantities used in the statements of our theoretical results.

A.1 Some calculations on multivariate Ornstein-Uhlenbeck processes

For convenience, we first collect some results about multivariate Ornstein-Uhlenbeck processes of the form (1), a detailed discussion of mvOU processes can be found in e.g. [50]. For a *time-invariant* process with drift matrix $\boldsymbol{A}$ and diffusion $\boldsymbol{\sigma}$, we have that

$$(\boldsymbol{X}_t | \boldsymbol{X}_0 = \boldsymbol{x}_0) \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t),$$

where

$$\boldsymbol{\mu}_t = e^{t\boldsymbol{A}}\boldsymbol{x}_0, \quad \boldsymbol{\Sigma}_t = \int_0^t e^{(t-\tau)\boldsymbol{A}}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top e^{(t-\tau)\boldsymbol{A}^\top} \mathrm{d}\tau. \tag{20}$$

If $\boldsymbol{X}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and writing $\boldsymbol{X}_{0t} = \boldsymbol{X}_0 \oplus \boldsymbol{X}_t$, some tedious but straightforward applications of conditional expectations and the tower property reveal that

$$\boldsymbol{X}_{0t} \sim \mathcal{N}(\boldsymbol{\mu}_{0t}, \boldsymbol{\Sigma}_{0t}), \quad \boldsymbol{\mu}_{0t} = \boldsymbol{\mu}_0 \oplus e^{t\boldsymbol{A}}\boldsymbol{\mu}_0, \quad \boldsymbol{\Sigma}_{0t} = \begin{bmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_0 e^{t\boldsymbol{A}^\top} \\ e^{t\boldsymbol{A}}\boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_t + e^{t\boldsymbol{A}}\boldsymbol{\Sigma}_0 e^{t\boldsymbol{A}^\top} \end{bmatrix}. \tag{21}$$

More generally, for a process $\mathrm{d}\boldsymbol{X}_t = (\boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{b})\mathrm{d}t + \boldsymbol{\sigma}\mathrm{d}\boldsymbol{B}_t$, one has:

$$\boldsymbol{\mu}_t = e^{t\boldsymbol{A}}\boldsymbol{x}_0 + (e^{t\boldsymbol{A}} - \mathbf{I})(\boldsymbol{A}^{-1}\boldsymbol{b}), \quad \boldsymbol{\Sigma}_t = \int_0^t e^{(t-s)\boldsymbol{A}}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top e^{(t-s)\boldsymbol{A}^\top} \mathrm{d}s. \tag{22}$$

The expressions for the covariance of the mvOU process can be obtained from the fact that the covariance evolves following a *Lyapunov equation*. For the case of constant coefficients, we state the following result.

Lyapunov equation solution It is easy to verify that $\dot{\boldsymbol{G}}_t = \boldsymbol{A}\boldsymbol{G}_t + \boldsymbol{G}_t\boldsymbol{A}^\top + \boldsymbol{Q}_t$ has solution

$$\boldsymbol{G}_t = \int_0^t e^{(t-s)\boldsymbol{A}}\boldsymbol{Q}_s e^{(t-s)\boldsymbol{A}^\top} \mathrm{d}s.$$

A.2 Bridges of multivariate Ornstein-Uhlenbeck processes

This problem has been studied by Chen et al. in [8], however the material therein is geared towards a control audience and considers a more general case where all coefficients are time-dependent. We will re-derive the results that we will need for processes of the form (1). Additionally, while in practice we typically use $\boldsymbol{\sigma} = \sigma\mathbf{I}$, we state some results in the general setting of non-isotropic noise.

Derivation of the dynamical OU-bridge (Theorem 1) Consider a mvOU process

$$\mathrm{d}\boldsymbol{X}_t = (\boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{b})\,\mathrm{d}t + \boldsymbol{\sigma}\,\mathrm{d}\boldsymbol{B}_t. \tag{23}$$

Now form the *controlled* version pinned at $(0, \boldsymbol{x}_0)$ and $(T, \boldsymbol{x}_T)$, where $\boldsymbol{u}_t(\boldsymbol{X}_t)$ is an additional force arising from the conditioning of the process at an endpoint:

$$\mathrm{d}\boldsymbol{X}_t = (A\boldsymbol{X}_t + \boldsymbol{b} + \boldsymbol{u}_t(\boldsymbol{X}_t))\,\mathrm{d}t + \boldsymbol{\sigma}\,\mathrm{d}\boldsymbol{B}_t. \tag{24}$$

The main result that we need is that $\boldsymbol{u}_t(\boldsymbol{x})$ is itself a linear time-dependent field whose coefficients are independent of $(\boldsymbol{x}_0, \boldsymbol{x}_T)$. In the above we consider the simplest case of classical optimal control, where the control is *directly* added to the system drift and there is no unobserved system.

Consider the system started at $(0, \boldsymbol{x}_0)$ and pinned to $(T, \boldsymbol{x}_T)$. For now, we make all statements for time-dependent coefficients and the diffusion is non-isotropic. We then form the Lagrangian:

$$\min_{\boldsymbol{u}} \mathbb{E}\left\{\int_0^T \|\boldsymbol{u}_t\|^2 \mathrm{d}t + \sup_{\lambda} \lambda\|\boldsymbol{X}_T - \boldsymbol{x}_T\|^2\right\} = \sup_{\lambda}\min_{\boldsymbol{u}} \mathbb{E}\left\{\int_0^T \|\boldsymbol{u}_t\|^2 \mathrm{d}t + \lambda\|\boldsymbol{X}_T - \boldsymbol{x}_T\|^2\right\}$$

Fixing some $\lambda > 0$, the resulting problem is known as a *linear-quadratic-Gaussian* problem [27] in the control literature:

$$\min_{\boldsymbol{u}} \mathbb{E}\left\{\frac{1}{2}\int_0^T \|\boldsymbol{u}_t\|^2 \mathrm{d}t + \frac{\lambda}{2}\|\boldsymbol{X}_T - \boldsymbol{x}_T\|^2\right\}, \tag{25}$$

for which we can write a HJB equation:

$$0 = \partial_t V_t(\boldsymbol{x}) + \min_{u}\left\{\widetilde{\mathcal{A}}_t[V_t](\boldsymbol{x}, \boldsymbol{u}_t(\boldsymbol{x})) + \frac{1}{2}\|\boldsymbol{u}_t(\boldsymbol{x})\|^2\right\}, \tag{26}$$

subject to the terminal boundary condition $V_T(\boldsymbol{x}) = \frac{\lambda}{2}\|\boldsymbol{x} - \boldsymbol{x}_T\|^2$, and the operators $\widetilde{\mathcal{A}}, \mathcal{A}$ are defined as

$$\widetilde{\mathcal{A}}_t[f] = \langle \boldsymbol{A}_t\boldsymbol{x} + \boldsymbol{b}_t + \boldsymbol{u}_t, \nabla_x f\rangle + \frac{1}{2}\nabla_x \cdot (\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^\top \nabla_x f), \tag{27}$$

$$\mathcal{A}_t[f] = \langle \boldsymbol{A}_t\boldsymbol{x} + \boldsymbol{b}_t, \nabla_x f\rangle + \frac{1}{2}\nabla_x \cdot (\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^\top \nabla_x f), \tag{28}$$

these are the generators of the controlled (24) and uncontrolled (23) SDEs respectively. Substituting all these in, we find that the HJBE is

$$0 = \partial_t V_t(\boldsymbol{x}) + \mathcal{A}_t[V_t](\boldsymbol{x}) + \min_{\boldsymbol{u}}\left[\langle \boldsymbol{u}_t, \nabla_x V_t\rangle + \frac{1}{2}\|\boldsymbol{u}_t\|^2\right]. \tag{29}$$

It follows from the first order condition on the "inner" problem that $\boldsymbol{u}_t = -\nabla_x V_t$, so

$$0 = \partial_t V_t(\boldsymbol{x}) + \mathcal{A}_t[V_t](\boldsymbol{x}) - \frac{1}{2}\|\nabla_x V_t(\boldsymbol{x})\|^2. \tag{30}$$

Use an ansatz that the value function is quadratic:

$$V_t(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{M}_t\boldsymbol{x} + \boldsymbol{c}_t^\top \boldsymbol{x} + d_t \Longrightarrow \nabla_x V_t(\boldsymbol{x}) = \boldsymbol{M}_t\boldsymbol{x} + \boldsymbol{c}_t. \tag{31}$$

Then:

$$\mathcal{A}_t[V_t] = \boldsymbol{x}^\top(\boldsymbol{A}_t^\top \boldsymbol{M}_t)\boldsymbol{x} + \langle \boldsymbol{A}_t^\top \boldsymbol{c}_t + \boldsymbol{M}_t\boldsymbol{b}_t, \boldsymbol{x}\rangle + \langle \boldsymbol{b}_t, \boldsymbol{c}_t\rangle + \frac{1}{2}\mathrm{tr}(\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^\top \boldsymbol{M}_t). \tag{32}$$

Now note that the quadratic form only depends on the *symmetric* part of the matrix, i.e.

$$\boldsymbol{x}^\top \boldsymbol{A}_t^\top \boldsymbol{M}_t\boldsymbol{x} = \frac{1}{2}\boldsymbol{x}^\top\left(\boldsymbol{A}_t^\top \boldsymbol{M}_t + \boldsymbol{M}_t\boldsymbol{A}_t\right)\boldsymbol{x}.$$

The HJBE is therefore

$$\begin{aligned}0 = &\frac{1}{2}\boldsymbol{x}^\top \dot{\boldsymbol{M}}_t\boldsymbol{x} + \dot{\boldsymbol{c}}_t^\top \boldsymbol{x} + \dot{d}_t \\ &+ \frac{1}{2}\boldsymbol{x}^\top(\boldsymbol{A}_t^\top \boldsymbol{M}_t + \boldsymbol{M}_t\boldsymbol{A}_t)\boldsymbol{x} + (\boldsymbol{A}_t^\top \boldsymbol{c}_t + \boldsymbol{M}_t^\top \boldsymbol{b}_t)^\top \boldsymbol{x} + \boldsymbol{b}_t^\top \boldsymbol{c}_t + \frac{1}{2}\mathrm{tr}(\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^\top \boldsymbol{M}_t) \\ &- \frac{1}{2}(\boldsymbol{M}_t\boldsymbol{x} + \boldsymbol{c}_t)^\top(\boldsymbol{M}_t\boldsymbol{x} + \boldsymbol{c}_t)\end{aligned} \tag{33}$$

So

$$0 = \dot{\boldsymbol{M}}_t + \boldsymbol{A}_t^\top \boldsymbol{M}_t + \boldsymbol{M}_t \boldsymbol{A}_t - \boldsymbol{M}_t^2, \tag{34}$$

$$0 = \dot{\boldsymbol{c}}_t + \boldsymbol{A}_t^\top \boldsymbol{c}_t + \boldsymbol{M}_t \boldsymbol{b}_t - \boldsymbol{M}_t \boldsymbol{c}_t. \tag{35}$$

with the boundary condition $\boldsymbol{M}_T = \lambda \mathbf{I}$ and $\boldsymbol{c}_t = -\lambda \boldsymbol{x}_T$.

Let $\boldsymbol{\Lambda}_t = \boldsymbol{M}_t^{-1}$, so that $\dot{\boldsymbol{\Lambda}}_t = -\boldsymbol{M}_t^{-1} \dot{\boldsymbol{M}}_t \boldsymbol{M}_t^{-1} = -\boldsymbol{\Lambda}_t \dot{\boldsymbol{M}}_t \boldsymbol{\Lambda}_t$. Substituting, we find that

$$\dot{\boldsymbol{\Lambda}}_t = \boldsymbol{\Lambda}_t \boldsymbol{A}_t^\top + \boldsymbol{A}_t \boldsymbol{\Lambda}_t - \mathbf{I}. \tag{36}$$

Actually we can rewrite the value function in a different way, which is more convenient for us:

$$V_t(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{k}_t)^\top \boldsymbol{M}_t (\boldsymbol{x} - \boldsymbol{k}_t) + const$$

in which case $\boldsymbol{c}_t = -\boldsymbol{M}_t \boldsymbol{k}_t$, and so we have $\boldsymbol{k}_T = \boldsymbol{x}_T$. The corresponding ODE for $\boldsymbol{k}_t$ is

$$\dot{\boldsymbol{k}}_t = \boldsymbol{A}_t \boldsymbol{k}_t + \boldsymbol{b}_t. \tag{37}$$

Let us now specialise to the case of time-invariant coefficients, which allows us to write down explicit expressions for the solutions:

$$\dot{\boldsymbol{k}}_t = \boldsymbol{A}\boldsymbol{k}_t + \boldsymbol{b} \Longrightarrow \boldsymbol{k}_t = e^{-(T-t)\boldsymbol{A}}(\boldsymbol{x}_T + \boldsymbol{A}^{-1}\boldsymbol{b}) - \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{38}$$

Rewriting the SDE drift to have the form $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{m})$ we have that $\boldsymbol{b} = -\boldsymbol{A}\boldsymbol{m} \Longrightarrow \boldsymbol{A}^{-1}\boldsymbol{b} = -\boldsymbol{m}$, when $\boldsymbol{A}$ is nonsingular. Then:

$$\boldsymbol{k}_t = e^{-(T-t)\boldsymbol{A}}(\boldsymbol{x}_T - \boldsymbol{m}) + \boldsymbol{m}. \tag{39}$$

Now we deal with the quadratic term $\boldsymbol{\Lambda}_t$. Let $\boldsymbol{G}_\tau = \boldsymbol{\Lambda}_{T-\tau}$, so that $\partial_\tau \boldsymbol{G}_\tau = -\dot{\boldsymbol{\Lambda}}_{T-\tau}$. Then $\boldsymbol{G}_\tau$ satisfies

$$\partial_\tau \boldsymbol{G}_\tau = -\boldsymbol{G}_\tau \boldsymbol{A}^\top - \boldsymbol{A}\boldsymbol{G}_\tau + \mathbf{I}, \quad \boldsymbol{G}_0 = 0. \tag{40}$$

So

$$\boldsymbol{G}_\tau = \int_0^\tau e^{-(\tau-s)\boldsymbol{A}} e^{-(\tau-s)\boldsymbol{A}^\top} \mathrm{d}s. \tag{41}$$

Substituting back, we find that

$$\boldsymbol{\Lambda}_t = \int_0^{T-t} e^{-(T-t-s)\boldsymbol{A}} e^{-(T-t-s)\boldsymbol{A}^\top} \mathrm{d}s = \int_0^{T-t} e^{-s\boldsymbol{A}} e^{-s\boldsymbol{A}^\top} \mathrm{d}s. \tag{42}$$

The bridge control is therefore

$$\boldsymbol{u}_{t|(\boldsymbol{x}_0, \boldsymbol{x}_T)} = -\boldsymbol{\Lambda}_t^{-1}(\boldsymbol{x} - \boldsymbol{k}_t) \tag{43}$$

For the general case of non-constant coefficients, we express them as solutions of ODEs with terminal boundary conditions.

$$\dot{\boldsymbol{k}}_t = \boldsymbol{A}_t \boldsymbol{k}_t + \boldsymbol{b}_t, \quad \boldsymbol{k}_T = \boldsymbol{x}_T, \tag{44}$$

$$\dot{\boldsymbol{\Lambda}}_t = \boldsymbol{\Lambda}_t \boldsymbol{A}_t^\top + \boldsymbol{A}_t \boldsymbol{\Lambda}_t - \boldsymbol{I}, \quad \boldsymbol{\Lambda}_T = 0. \tag{45}$$

In practice, both of these equations can be approximately solved by numerical integration in time.

Finally we remark here that the diffusion component does not play a role in the control, which is classical.

"Static" Ornstein-Uhlenbeck bridge statistics (Theorem 2) As was studied by [8] for a more general scenario of time-varying processes, the mvOU process and its conditioned versions are Gaussian processes. Under these assumptions, $(\boldsymbol{X}_s, \boldsymbol{X}_t)$ has joint mean

$$\boldsymbol{\mu}_s \oplus \boldsymbol{\mu}_t, \tag{46}$$

and covariance

$$\begin{bmatrix} \boldsymbol{\Phi}_s & \boldsymbol{\Phi}_s e^{(t-s)\boldsymbol{A}^\top} \\ e^{(t-s)\boldsymbol{A}} \boldsymbol{\Phi}_s & \boldsymbol{\Phi}_t \end{bmatrix}. \tag{47}$$

where $\boldsymbol{\Phi}_t$ is the covariance of the process started from a point mass at time t:

$$\boldsymbol{\Phi}_t = \int_0^t e^{(t-s)\boldsymbol{A}}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top e^{(t-s)\boldsymbol{A}^\top}\mathrm{d}s, \quad \boldsymbol{\Phi}_t = \boldsymbol{\Phi}_{t-s} + e^{(t-s)\boldsymbol{A}}\boldsymbol{\Phi}_s e^{(t-s)\boldsymbol{A}^\top}. \tag{48}$$

By taking the Schur complement, we have that $\boldsymbol{X}_s|\boldsymbol{X}_t$ is distributed with variance

$$\mathbb{V}[\boldsymbol{X}_s|\boldsymbol{X}_t] = \boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s e^{(t-s)\boldsymbol{A}^\top}\boldsymbol{\Phi}_t^{-1}e^{(t-s)\boldsymbol{A}}\boldsymbol{\Phi}_s \tag{49}$$

and mean

$$\mathbb{E}[\boldsymbol{X}_s|\boldsymbol{X}_t] = \boldsymbol{\mu}_s + \boldsymbol{\Phi}_s e^{(t-s)\boldsymbol{A}^\top}\boldsymbol{\Phi}_t^{-1}(\boldsymbol{X}_t - \boldsymbol{\mu}_t). \tag{50}$$

More generally, we can consider the three-point correlation $(\boldsymbol{X}_r, \boldsymbol{X}_s, \boldsymbol{X}_t)$. Then the mean is $\boldsymbol{\mu}_r \oplus \boldsymbol{\mu}_s \oplus \boldsymbol{\mu}_t$ and the covariance is

$$\begin{bmatrix} \boldsymbol{\Phi}_r & \boldsymbol{\Phi}_r e^{(s-r)\boldsymbol{A}^\top} & \boldsymbol{\Phi}_r e^{(t-r)\boldsymbol{A}^\top} \\ \cdot & \boldsymbol{\Phi}_s & \boldsymbol{\Phi}_s e^{(t-s)\boldsymbol{A}^\top} \\ \cdot & \cdot & \boldsymbol{\Phi}_t \end{bmatrix} \tag{51}$$

Using the Schur complement again, we have that $(\boldsymbol{X}_r, \boldsymbol{X}_s)|\boldsymbol{X}_t$ has covariance

$$\mathbb{V}[\boldsymbol{X}_r, \boldsymbol{X}_s|\boldsymbol{X}_t] = \boldsymbol{\Phi}_r e^{(s-r)\boldsymbol{A}^\top} - \boldsymbol{\Phi}_r e^{(t-r)\boldsymbol{A}^\top}\boldsymbol{\Phi}_t^{-1}e^{(t-s)\boldsymbol{A}}\boldsymbol{\Phi}_s. \tag{52}$$

For the case of time-varying coefficients, we need to introduce the state transition matrix $\boldsymbol{\Psi}_{ts}$ [27] which describes the deterministic aspects of evolution between $s < t$ under $(\boldsymbol{A}_t)_t$. That is, for a dynamics $\dot{\boldsymbol{x}}_t = \boldsymbol{A}_t\boldsymbol{x}_t$ one has $\boldsymbol{x}_t = \boldsymbol{\Psi}_{t,s}\boldsymbol{x}_s$ and $\boldsymbol{\Psi}(t,s) = \boldsymbol{\Psi}(t,r)\boldsymbol{\Psi}(r,s)$ for $t \geqslant r \geqslant s$. In this case there is not an explicit expression for $\boldsymbol{\Phi}_t$. Instead, let $(\boldsymbol{\Phi}_t)_t$ be the unconditional variance evolutions for such a process started at $t = 0$, obtained by solving

$$\dot{\boldsymbol{\Phi}}_t = \boldsymbol{A}_t\boldsymbol{\Phi}_t + \boldsymbol{\Phi}_t\boldsymbol{A}_t^\top + \boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^\top, \quad \boldsymbol{\Phi}_0 = \boldsymbol{0}. \tag{53}$$

Then the more general result from [8] for the covariance of $(\boldsymbol{X}_r, \boldsymbol{X}_s, \boldsymbol{X}_r)$ is

$$\begin{bmatrix} \boldsymbol{\Phi}_r & \boldsymbol{\Phi}_r\boldsymbol{\Psi}_{sr}^\top & \boldsymbol{\Phi}_r\boldsymbol{\Psi}_{tr}^\top \\ \boldsymbol{\Psi}_{sr}\boldsymbol{\Phi}_r & \boldsymbol{\Phi}_s & \boldsymbol{\Phi}_s\boldsymbol{\Psi}_{ts}^\top \\ \boldsymbol{\Psi}_{tr}\boldsymbol{\Phi}_r & \boldsymbol{\Psi}_{ts}\boldsymbol{\Phi}_s & \boldsymbol{\Phi}_t. \end{bmatrix} \tag{54}$$

From this result, the same Schur complement computation yields expressions for the conditional covariances and mean:

$$\mathbb{V}[\boldsymbol{X}_r, \boldsymbol{X}_s|\boldsymbol{X}_t] = \boldsymbol{\Phi}_r\boldsymbol{\Psi}_{sr}^\top - \boldsymbol{\Phi}_r\boldsymbol{\Psi}_{tr}^\top\boldsymbol{\Phi}_t^{-1}\boldsymbol{\Psi}_{ts}\boldsymbol{\Phi}_s, \tag{55}$$

$$\mathbb{V}[\boldsymbol{X}_s|\boldsymbol{X}_t] = \boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Psi}_{ts}^\top\boldsymbol{\Phi}_t^{-1}\boldsymbol{\Psi}_{ts}\boldsymbol{\Phi}_s, \tag{56}$$

$$\mathbb{E}[\boldsymbol{X}_s|\boldsymbol{X}_t] = \boldsymbol{\mu}_s + \boldsymbol{\Phi}_s\boldsymbol{\Psi}_{ts}^\top\boldsymbol{\Phi}_t^{-1}(\boldsymbol{X}_t - \boldsymbol{\mu}_t) \tag{57}$$

A.3 Simulation-free Schrödinger bridges with linear reference dynamics

Having characterised the solution of the SBP for general reference processes and now that we have derived the score and flow for the Ornstein-Uhlenbeck bridge, the path is clear towards a simulation-free scheme for learning Schrödinger bridges where the reference dynamics are given by a *linear* SDE.

For solution of the static SBP problem (18) (equivalently, (SBP-static)), the transition kernel of the reference dynamics is given by

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{Q}_{0t}}{\mathrm{d}\boldsymbol{x}\otimes\mathrm{d}\boldsymbol{x}}(\boldsymbol{x}_0, \boldsymbol{x}_t) &= \frac{1}{Z_t}\exp\left(-\frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{\mu}_t^{\boldsymbol{x}_0})^\top\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_t^{\boldsymbol{x}_0})\right)\frac{\mathrm{d}\boldsymbol{Q}_0}{\mathrm{d}\boldsymbol{x}}(\boldsymbol{x}_0) \\ \implies -\log\left(\frac{\mathrm{d}\boldsymbol{Q}_{0t}}{\mathrm{d}\boldsymbol{x}\otimes\mathrm{d}\boldsymbol{x}}\right) &= \frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{\mu}_t^{\boldsymbol{x}_0})^\top\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_t^{\boldsymbol{x}_0}) + \log Z_t - \log\left(\frac{\mathrm{d}\boldsymbol{Q}_0}{\mathrm{d}\boldsymbol{x}}(\boldsymbol{x}_0)\right), \end{aligned} \tag{58}$$

where $\boldsymbol{\mu}_t^{\boldsymbol{x}_0}$ denotes the mean at time t conditional on $(0, \boldsymbol{x}_0)$. Notably, the last two terms depend only on $\boldsymbol{x}_0$. It is a classical result (and very easy to show) that these kinds of terms do not affect the minimiser of (18) and so they are immaterial. The cost function to use is thus effectively

$$C(\boldsymbol{x}_0, \boldsymbol{x}_t) = \frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{\mu}_t^{\boldsymbol{x}_0})^\top\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_t^{\boldsymbol{x}_0}). \tag{59}$$

Once a solution $\boldsymbol{\pi}$ of the static SBP is on hand, we want to utilise the stochastic regression objective on the conditional score and flow. Sampling $(\boldsymbol{x}_0, \boldsymbol{x}_t)$ from $\boldsymbol{\pi}$, recall that $\boldsymbol{P}^{(x_0,x_1)} = \boldsymbol{Q}^{(x_0,x_1)}$ so we want to sample from the $\boldsymbol{Q}$-bridge using (11), (12):

$$t \sim U[0,1], \quad \boldsymbol{x}_t^{(x_0,x_1)} \sim \mathcal{N}(\boldsymbol{\mu}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_1)}, \boldsymbol{\Sigma}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_1)}). \tag{60}$$

Equations (9) and (12) give us the score and flow at $\boldsymbol{x}_t$.

A.4 Characterisation of the $\mathbb{Q}$-GSB (Theorem 3)

We will construct the $\mathbb{Q}$-GSB utilising its characterisation (2). Our approach is similar to that of [5] – we first obtain explicit formulae for the static $\mathbb{Q}$-GSB (SBP-static) and then build towards the dynamical $\mathbb{Q}$-GSB (SBP-dyn) by using the characterisation of $\mathbb{Q}$-bridges.

Static $\mathbb{Q}$-GSB The standard Gaussian EOT problem has a well known solution [30, 20, 5, 21]. Here we use the notations of [20] and define

$$\mathrm{OT}^{\otimes}_{\sigma^2}(\alpha,\beta) := \min_{\pi\in\Pi(\alpha,\beta)} \frac{1}{2}\int \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \,\mathrm{d}\pi(\boldsymbol{x},\boldsymbol{y}) + \sigma^2\,\mathrm{H}(\pi|\alpha\otimes\beta), \tag{61}$$

where $\sigma > 0$ is the regularisation level and $\alpha = \mathcal{N}(\boldsymbol{a},\boldsymbol{A})$ and $\beta = \mathcal{N}(\boldsymbol{b},\boldsymbol{B})$. The solution to (61) in the Gaussian case is given by [20, Theorem 1]

$$\mathrm{OT}^{\otimes}_{\sigma^2}(\alpha,\beta) = \frac{1}{2}\|\boldsymbol{a}-\boldsymbol{b}\|_2^2 + \frac{1}{2}\mathsf{B}^{\otimes}_{\sigma^2}(\boldsymbol{A},\boldsymbol{B}) \tag{62}$$

$$\mathsf{B}^{\otimes}_{\sigma^2}(\boldsymbol{A},\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B}) - 2\mathrm{tr}(\boldsymbol{C}) + \sigma^2\log\det\left(\frac{1}{\sigma^2}\boldsymbol{C}+\boldsymbol{I}\right), \tag{63}$$

$$\boldsymbol{C} = \boldsymbol{A}^{1/2}\left(\boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2} + \frac{\sigma^4}{4}\boldsymbol{I}\right)^{1/2}\boldsymbol{A}^{-1/2} - \frac{\sigma^2}{2}\boldsymbol{I}. \tag{64}$$

We are, of course, interested in a slightly different problem, namely, for $\rho_0 = \mathcal{N}(\boldsymbol{a},\boldsymbol{\mathcal{A}})$ and $\rho_T = \mathcal{N}(\boldsymbol{b},\boldsymbol{\mathcal{B}})$, we seek to solve $\min_{\pi\in\Pi(\rho_0,\rho_T)} \mathrm{KL}(\pi|\mathbb{Q}_{0,T})$. Expanding the definition of KL and simplifying, we get

$$\begin{aligned}
(\text{SBP-static}) &= \min_{\pi\in\Pi(\rho_0,\rho_T)} \int \mathrm{d}\pi \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\mathbb{Q}_{0,T}}\right) && (65)\\
&= \min_{\pi\in\Pi(\rho_0,\rho_T)} -\int \mathrm{d}\pi \log\left(\frac{\mathrm{d}\mathbb{Q}_{0,T}}{\mathrm{d}x_0\otimes\mathrm{d}x_T}\right) + \mathrm{H}(\pi|\rho_0\otimes\rho_T) + \mathrm{H}(\rho_0|\mathrm{d}x_0) + \mathrm{H}(\rho_T|\mathrm{d}x_T) && (66)\\
&\simeq \min_{\pi\in\Pi(\rho_0,\rho_T)} -\int \mathrm{d}\pi \log\left(\frac{\mathrm{d}\mathbb{Q}_{0,T}}{\mathrm{d}x_0\otimes\mathrm{d}x_T}\right) + \mathrm{H}(\pi|\rho_0\otimes\rho_T) && (67)\\
&\simeq \min_{\pi\in\Pi(\rho_0,\rho_T)} -\int \mathrm{d}\pi(x_0,x_T) \log\left(\frac{\mathrm{d}\mathbb{Q}_{T|0}(x_T|x_0)}{\mathrm{d}x_T}\right) + \mathrm{H}(\pi|\rho_0\otimes\rho_T) && (68)
\end{aligned}$$

where by $\simeq$ we denote equality of the objective up to additive constants which do not affect the minimiser π. We note that the last line is exactly (61) with $\alpha,\beta = \rho_0,\rho_T$ and $\sigma = 1$ and a modified cost. Further, recognise $\mathrm{d}\mathbb{Q}_{T|0}(x_T|x_0)/\mathrm{d}x_T$ as the $\mathbb{Q}$-transition density, and $\mathbb{Q}_{T|0} = \mathcal{N}(\boldsymbol{\mu}_T^{\boldsymbol{x}_0}, \boldsymbol{\Sigma}_T)$. Thus,

$$(\text{SBP-static}) \simeq \min_{\pi\in\Pi(\rho_0,\rho_T)} \frac{1}{2}\int (\boldsymbol{x}_T - \boldsymbol{\mu}_T^{\boldsymbol{x}_0})^\top \boldsymbol{\Sigma}_T^{-1}(\boldsymbol{x}_T - \boldsymbol{\mu}_T^{\boldsymbol{x}_0})\,\mathrm{d}\pi(\boldsymbol{x}_0,\boldsymbol{x}_T) + \mathrm{H}(\pi|\rho_0\otimes\rho_T) \tag{69}$$

where we remind that $\boldsymbol{\mu}_t^{\boldsymbol{x}_0}$ is the mean at time t conditional on starting at $(0,\boldsymbol{x}_0)$, and $\boldsymbol{\Sigma}_t$ is the covariance at time t, started at a point mass, i.e. $\boldsymbol{\Sigma}_0 = 0$.

In particular, we have $\boldsymbol{\Sigma}_t = \boldsymbol{\Phi}_t$, using the notation of Theorem 2. Recall that in the case of constant coefficients, we have the following explicit expressions:

$$\boldsymbol{\mu}_t^{\boldsymbol{x}_0} = e^{t\boldsymbol{A}}(\boldsymbol{x}_0 - \boldsymbol{m}) + \boldsymbol{m}, \tag{70}$$

$$\boldsymbol{\Phi}_t = \int_0^t e^{(t-s)\boldsymbol{A}}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top e^{(t-s)\boldsymbol{A}^\top}\mathrm{d}s. \tag{71}$$

Define the following changes of variable $(\boldsymbol{x}_0, \boldsymbol{x}_T) \mapsto (\overline{\boldsymbol{x}}_0, \overline{\boldsymbol{x}}_T)$, injective whenever $\boldsymbol{\Sigma}_T$ is positive definite:

$$\begin{aligned}
\overline{\boldsymbol{x}}_0 &= T_0(\boldsymbol{x}_0) = \boldsymbol{\Sigma}_T^{-1/2} \boldsymbol{\mu}_T^{\boldsymbol{x}_0}, \\
\overline{\boldsymbol{x}}_T &= T_T(\boldsymbol{x}_T) = \boldsymbol{\Sigma}_T^{-1/2} \boldsymbol{x}_T
\end{aligned}$$

Let $\overline{\rho}_0 = (T_0)_{\#}\rho_0$ and $\overline{\rho}_T = (T_T)_{\#}\rho_T$, and $\overline{\pi} = (T_0 \times T_T)_{\#}\pi$ and note that the relative entropy is invariant to this change of coordinates. So the problem (69) is equivalent to

$$\min_{\overline{\pi} \in \Pi(\overline{\rho}_0, \overline{\rho}_T)} \frac{1}{2} \int \|\overline{\boldsymbol{x}}_T - \overline{\boldsymbol{x}}_0\|_2^2 \, \mathrm{d}\overline{\pi}(\overline{\boldsymbol{x}}_0, \overline{\boldsymbol{x}}_T) + \mathrm{H}(\overline{\pi} | \overline{\rho}_0 \otimes \overline{\rho}_T). \tag{72}$$

This is exactly $\mathrm{OT}^{\otimes}_{\sigma^2=1}(\overline{\rho}_0, \overline{\rho}_T)$ as per (61) with $\overline{\rho}_0 = \mathcal{N}(\overline{\boldsymbol{a}}, \overline{\boldsymbol{\mathcal{A}}})$ and $\overline{\rho}_T = \mathcal{N}(\overline{\boldsymbol{b}}, \overline{\boldsymbol{\mathcal{B}}})$. The transformed means and covariances are

$$\begin{aligned}
\overline{\boldsymbol{a}} &= \boldsymbol{\Sigma}_T^{-1/2}(e^{T\boldsymbol{A}}(\boldsymbol{a} - \boldsymbol{m}) + \boldsymbol{m}) \\
\overline{\boldsymbol{\mathcal{A}}} &= \boldsymbol{\Sigma}_T^{-1/2} e^{T\boldsymbol{A}} \boldsymbol{\mathcal{A}} e^{T\boldsymbol{A}^\top} \boldsymbol{\Sigma}_T^{-1/2} \\
\overline{\boldsymbol{b}} &= \boldsymbol{\Sigma}_T^{-1/2} \boldsymbol{b} \\
\overline{\boldsymbol{\mathcal{B}}} &= \boldsymbol{\Sigma}_T^{-1/2} \boldsymbol{\mathcal{B}} \boldsymbol{\Sigma}_T^{-1/2}
\end{aligned}$$

We remark that for time-dependent coefficients our approach still applies, however computation of the transformation coefficients will be in terms of the general state transition matrix instead of matrix exponentials. In what follows, we focus on results for constant coefficients for simplicity and their practical relevance.

The solution to (72) is therefore

$$(\overline{\boldsymbol{x}}_0, \overline{\boldsymbol{x}}_T) \sim \overline{\pi} = \mathcal{N}\left(\begin{bmatrix} \overline{\boldsymbol{a}} \\ \overline{\boldsymbol{b}} \end{bmatrix}, \begin{bmatrix} \overline{\boldsymbol{\mathcal{A}}} & \overline{\boldsymbol{\mathcal{C}}} \\ \overline{\boldsymbol{\mathcal{C}}}^\top & \overline{\boldsymbol{\mathcal{B}}} \end{bmatrix} \right),$$

where $\overline{\boldsymbol{\mathcal{C}}}$ is the cross-covariance the transport plan, given by (64).

Dynamic $\mathbb{Q}$-GSB : marginals and covariance Let $\boldsymbol{X}_t | (\boldsymbol{x}_0, \boldsymbol{x}_T)$ denote the $\mathbb{Q}$-bridge pinned at $(\boldsymbol{x}_0, \boldsymbol{x}_T)$. Then by Theorem 2,

$$\boldsymbol{X}_t | (\boldsymbol{x}_0, \boldsymbol{x}_T) = \mathcal{N}(\boldsymbol{\mu}_{t|(\boldsymbol{x}_0, \boldsymbol{x}_T)}, \boldsymbol{\Sigma}_{t|(\boldsymbol{x}_0, \boldsymbol{x}_T)}) = \mathcal{N}(\boldsymbol{\mu}_{t|(\boldsymbol{x}_0, \boldsymbol{x}_T)}, \boldsymbol{\Omega}_t).$$

Invoking the inverse mappings T_0^{-1}, T_T^{-1} and substituting the expression for $\boldsymbol{\mu}_{t|(\boldsymbol{x}_0, \boldsymbol{x}_T)}$ from 2, we get

$$\begin{aligned}
\boldsymbol{\mu}_{t|(T_0^{-1}(\overline{\boldsymbol{x}}_0), T_T^{-1}(\overline{\boldsymbol{x}}_T))} &= \underbrace{\left(e^{-(T-t)\boldsymbol{A}} - \boldsymbol{\Gamma}_t\right) \boldsymbol{\Sigma}_T^{1/2}}_{\mathfrak{A}_t} \overline{\boldsymbol{x}}_0 + \underbrace{\boldsymbol{\Gamma}_t \boldsymbol{\Sigma}_T^{1/2}}_{\mathfrak{B}_t} \overline{\boldsymbol{x}}_T + \underbrace{\left(\boldsymbol{I} - e^{-(T-t)\boldsymbol{A}}\right) \boldsymbol{m}}_{\mathfrak{c}_t} && (73) \\
&= \mathfrak{A}_t \overline{\boldsymbol{x}}_0 + \mathfrak{B}_t \overline{\boldsymbol{x}}_T + \mathfrak{c}_t. && (74)
\end{aligned}$$

where $\boldsymbol{\Gamma}_t = \boldsymbol{\Phi}_t e^{(T-t)\boldsymbol{A}^\top} \boldsymbol{\Phi}_T^{-1}$. For clarity, we state first a general formula:

Lemma 1. *Let* $(\boldsymbol{X}_0, \boldsymbol{X}_1) \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{\mu}_1 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{10} & \boldsymbol{\Sigma}_{11} \end{bmatrix} \right)$. *Let*

$$\boldsymbol{Y}_{\boldsymbol{X}_0, \boldsymbol{X}_1} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{X}_0 + \boldsymbol{B}\boldsymbol{X}_1 + \boldsymbol{c}, \boldsymbol{\Omega}).$$

Then, $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{A}\boldsymbol{\mu}_0 + \boldsymbol{B}\boldsymbol{\mu}_1 + \boldsymbol{c}$ *and* $\mathbb{V}[\boldsymbol{Y}] = \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{\Sigma}_{00}\boldsymbol{A}^\top + \boldsymbol{B}\boldsymbol{\Sigma}_{11}\boldsymbol{B}^\top + \boldsymbol{A}\boldsymbol{\Sigma}_{01}\boldsymbol{B}^\top + \boldsymbol{B}\boldsymbol{\Sigma}_{01}^\top\boldsymbol{A}^\top$.

Application to (74) gives us the following formula for the variance of the bridge at time t:

$$\mathbb{V}[\boldsymbol{X}_t] = \boldsymbol{\Omega}_t + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{A}}} \mathfrak{A}_t^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{B}}} \mathfrak{B}_t^\top + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{C}}} \mathfrak{B}_t^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{C}}}^\top \mathfrak{A}_t^\top =: \boldsymbol{\Xi}_t. \tag{75}$$

Substituting and simplifying, we get the following expressions for each of the terms:

$$\begin{aligned}
\mathfrak{A}_t \overline{\boldsymbol{\mathcal{A}}} \mathfrak{A}_t^\top &= (\boldsymbol{I} - \boldsymbol{\Gamma}_t e^{(T-t)\boldsymbol{A}}) e^{t\boldsymbol{A}} \boldsymbol{\mathcal{A}} e^{t\boldsymbol{A}^\top} (\boldsymbol{I} - \boldsymbol{\Gamma}_t e^{(T-t)\boldsymbol{A}})^\top, && (76) \\
\mathfrak{B}_t \overline{\boldsymbol{\mathcal{B}}} \mathfrak{B}_t^\top &= \boldsymbol{\Gamma}_t \boldsymbol{\mathcal{B}} \boldsymbol{\Gamma}_t^\top, && (77) \\
\mathfrak{A}_t \overline{\boldsymbol{\mathcal{C}}} \mathfrak{B}_t^\top &= (e^{-(T-t)\boldsymbol{A}} - \boldsymbol{\Gamma}_t) \boldsymbol{\Sigma}_T^{1/2} \overline{\boldsymbol{\mathcal{C}}} \boldsymbol{\Sigma}_T^{1/2} \boldsymbol{\Gamma}_t^\top. && (78)
\end{aligned}$$

Similarly, the mean can be computed as

$$\mathbb{E}[\boldsymbol{X}_t] = \mathfrak{A}_t \overline{\boldsymbol{a}} + \mathfrak{B}_t \overline{\boldsymbol{b}} + \mathfrak{c}_t \tag{79}$$

$$= \left(e^{-(T-t)\boldsymbol{A}} - \boldsymbol{\Gamma}_t\right) e^{T\boldsymbol{A}}(\boldsymbol{a} - \boldsymbol{m}) + \boldsymbol{\Gamma}_t(\boldsymbol{b} - \boldsymbol{m}) + \boldsymbol{m} =: \boldsymbol{\nu}_t. \tag{80}$$

To calculate the covariance of the SB process, we use again the disintegration property (2). We have from (52) for the $\mathbb{Q}$-bridges and $0 < s < t < T$:

$$\mathbb{V}[\boldsymbol{X}_s, \boldsymbol{X}_t | \boldsymbol{X}_0, \boldsymbol{X}_T] = \boldsymbol{\Phi}_s e^{(t-s)\boldsymbol{A}^\top} - \boldsymbol{\Phi}_s e^{(T-s)\boldsymbol{A}^\top} \boldsymbol{\Phi}_T^{-1} e^{(T-t)\boldsymbol{A}} \boldsymbol{\Phi}_t =: \boldsymbol{\Omega}_{s,t}. \tag{81}$$

Now,

$$\mathbb{V}[\boldsymbol{X}_s, \boldsymbol{X}_t] = \mathbb{E}[(\boldsymbol{X}_s - \boldsymbol{\nu}_s)(\boldsymbol{X}_t - \boldsymbol{\nu}_t)^\top] \tag{82}$$

$$= \mathbb{E}[\boldsymbol{X}_s \boldsymbol{X}_t^\top] - \boldsymbol{\nu}_s \boldsymbol{\nu}_t^\top \tag{83}$$

$$= \mathbb{E}[\mathbb{E}[\boldsymbol{X}_s \boldsymbol{X}_t^\top | \boldsymbol{X}_0, \boldsymbol{X}_T]] - \boldsymbol{\nu}_s \boldsymbol{\nu}_t^\top. \tag{84}$$

and

$$\mathbb{E}[\boldsymbol{X}_s \boldsymbol{X}_t^\top | \boldsymbol{X}_0, \boldsymbol{X}_T] = \mathbb{V}[\boldsymbol{X}_s, \boldsymbol{X}_t | \boldsymbol{X}_0, \boldsymbol{X}_T] + \boldsymbol{\mu}_{s|(\boldsymbol{X}_0, \boldsymbol{X}_T)} \boldsymbol{\mu}_{t|(\boldsymbol{X}_0, \boldsymbol{X}_T)}^\top, \tag{85}$$

in which the first (variance) term doesn't actually depend on $(\boldsymbol{X}_0, \boldsymbol{X}_T)$. So,

$$\mathbb{E}[\boldsymbol{X}_s \boldsymbol{X}_t^\top] = \boldsymbol{\Omega}_{st} + \mathbb{E}_{(\boldsymbol{X}_0, \boldsymbol{X}_T)}[\boldsymbol{\mu}_{s|(\boldsymbol{X}_0, \boldsymbol{X}_T)} \boldsymbol{\mu}_{t|(\boldsymbol{X}_0, \boldsymbol{X}_T)}^\top]. \tag{86}$$

In fact let's switch to the "mapped" coordinates $\overline{\boldsymbol{X}}_0, \overline{\boldsymbol{X}}_T$ for the endpoints. We abuse notation and omit the inverse map in what follows. Then, from previously,

$$\boldsymbol{\mu}_{t|(\overline{\boldsymbol{X}}_0, \overline{\boldsymbol{X}}_T)} = \mathfrak{A}_t \overline{\boldsymbol{X}}_0 + \mathfrak{B}_t \overline{\boldsymbol{X}}_T + \mathfrak{c}_t. \tag{87}$$

Expanding, collecting and cancelling terms, we have that

$$\mathbb{E}[\boldsymbol{\mu}_{s|(\overline{\boldsymbol{X}}_0, \overline{\boldsymbol{X}}_T)} \boldsymbol{\mu}_{t|(\overline{\boldsymbol{X}}_0, \overline{\boldsymbol{X}}_T)}^\top] - \boldsymbol{\nu}_s \boldsymbol{\nu}_t^\top = \mathfrak{A}_s \mathbb{V}[\overline{\boldsymbol{X}}_0] \mathfrak{A}_t^\top + \mathfrak{A}_s \mathbb{V}[\overline{\boldsymbol{X}}_0, \overline{\boldsymbol{X}}_T] \mathfrak{B}_t^\top + \mathfrak{B}_s \mathbb{V}[\overline{\boldsymbol{X}}_T, \overline{\boldsymbol{X}}_0] \mathfrak{A}_t^\top + \mathfrak{B}_s \mathbb{V}[\overline{\boldsymbol{X}}_T] \mathfrak{B}_t^\top \tag{88}$$

$$= \mathfrak{A}_s \overline{\boldsymbol{\mathcal{A}}} \mathfrak{A}_t^\top + \mathfrak{A}_s \overline{\boldsymbol{\mathcal{C}}} \mathfrak{B}_t^\top + \mathfrak{B}_s \overline{\boldsymbol{\mathcal{C}}}^\top \mathfrak{A}_t^\top + \mathfrak{B}_s \overline{\boldsymbol{\mathcal{B}}} \mathfrak{B}_t^\top. \tag{89}$$

Putting everything together, we get

$$\mathbb{V}[\boldsymbol{X}_s, \boldsymbol{X}_t] = \boldsymbol{\Omega}_{st} + \mathfrak{A}_s \overline{\boldsymbol{\mathcal{A}}} \mathfrak{A}_t^\top + \mathfrak{A}_s \overline{\boldsymbol{\mathcal{C}}} \mathfrak{B}_t^\top + \mathfrak{B}_s \overline{\boldsymbol{\mathcal{C}}}^\top \mathfrak{A}_t^\top + \mathfrak{B}_s \overline{\boldsymbol{\mathcal{B}}} \mathfrak{B}_t^\top =: \boldsymbol{\Xi}_{s,t}. \tag{90}$$

Dynamic $\mathbb{Q}$-GSB : SDE representation We proceed via the generator route also used by [5]. Expanding the covariance (90), we find that

$$\mathbb{V}[\boldsymbol{X}_t, \boldsymbol{X}_{t+h}] = \boldsymbol{\Xi}_{t,t+h} = \boldsymbol{\Omega}_{t,t+h} + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{A}}} \mathfrak{A}_{t+h}^\top + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{C}}} \mathfrak{B}_{t+h}^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{C}}}^\top \mathfrak{A}_{t+h}^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{B}}} \mathfrak{B}_{t+h}^\top \tag{91}$$

$$= \boldsymbol{\Xi}_t + h\left\{(\partial_{t'} \boldsymbol{\Omega}_{t,t'})(t) + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{A}}} \dot{\mathfrak{A}}_t^\top + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{C}}} \dot{\mathfrak{B}}_t^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{C}}}^\top \dot{\mathfrak{A}}_t^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{B}}} \dot{\mathfrak{B}}_t^\top\right\} + \ldots \tag{92}$$

$$= \boldsymbol{\Xi}_t + h\boldsymbol{S}_t + \ldots \tag{93}$$

where we have set the key quantity

$$\boldsymbol{S}_t = (\partial_{t'} \boldsymbol{\Omega}_{t,t'})(t) + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{A}}} \dot{\mathfrak{A}}_t^\top + \mathfrak{A}_t \overline{\boldsymbol{\mathcal{C}}} \dot{\mathfrak{B}}_t^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{C}}}^\top \dot{\mathfrak{A}}_t^\top + \mathfrak{B}_t \overline{\boldsymbol{\mathcal{B}}} \dot{\mathfrak{B}}_t^\top \tag{94}$$

Now, $(\boldsymbol{X}_{t+h}, \boldsymbol{X}_t)$ are jointly Gaussian with mean and covariance

$$\begin{bmatrix} \boldsymbol{\nu}_{t+h} \\ \boldsymbol{\nu}_t \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Xi}_{t+h} & \boldsymbol{\Xi}_{t+h,t} \\ \boldsymbol{\Xi}_{t,t+h} & \boldsymbol{\Xi}_t \end{bmatrix}. \tag{95}$$

So

$$\mathbb{E}[\boldsymbol{X}_{t+h} | \boldsymbol{X}_t] = \boldsymbol{\nu}_{t+h} + \boldsymbol{\Xi}_{t,t+h}^\top \boldsymbol{\Xi}_t^{-1}(\boldsymbol{X}_t - \boldsymbol{\nu}_t) \tag{96}$$

$$= \boldsymbol{X}_t + h\left[\dot{\boldsymbol{\nu}}_t + \boldsymbol{S}_t^\top \boldsymbol{\Xi}_t^{-1}(\boldsymbol{X}_t - \boldsymbol{\nu}_t)\right] \tag{97}$$

$$\mathbb{V}[\boldsymbol{X}_{t+h} | \boldsymbol{X}_t] = \boldsymbol{\Xi}_{t+h} - \boldsymbol{\Xi}_{t,t+h}^\top \boldsymbol{\Xi}_t^{-1} \boldsymbol{\Xi}_{t,t+h} \tag{98}$$

$$= h(\dot{\boldsymbol{\Sigma}}_t - (\boldsymbol{S}_t + \boldsymbol{S}_t^\top)). \tag{99}$$

Let $\boldsymbol{x} \mapsto u(\boldsymbol{x})$ be a twice differentiable test function. Then:

$$\mathbb{E}\left[u(\boldsymbol{X}_{t+h})|\boldsymbol{X}_t = \boldsymbol{x}\right] = \mathbb{E}\left[u(\boldsymbol{X}_t) + (\partial_x u(\boldsymbol{X}_t))^\top(\boldsymbol{X}_{t+h} - \boldsymbol{X}_t) + \frac{1}{2}(\boldsymbol{X}_{t+h} - \boldsymbol{X}_t)^\top(\partial^2_{xx}u(\boldsymbol{X}_t))(\boldsymbol{X}_{t+h} - \boldsymbol{X}_t)\middle|\boldsymbol{X}_t = \boldsymbol{x}\right] \tag{100}$$

$$= u(\boldsymbol{x}) + h(\partial_x u(\boldsymbol{x}))^\top\left(\dot{\boldsymbol{\nu}}_t + \boldsymbol{S}_t^\top \boldsymbol{\Xi}_t^{-1}(\boldsymbol{x} - \boldsymbol{\nu}_t)\right) + \frac{1}{2}\mathbb{E}\left[(\boldsymbol{X}_{t+h} - \boldsymbol{X}_t)^\top(\partial^2_{xx}u(\boldsymbol{X}_t))(\boldsymbol{X}_{t+h} - \boldsymbol{X}_t)\middle|\boldsymbol{X}_t = \boldsymbol{x}\right] \tag{101}$$

$$= u(\boldsymbol{x}) + h(\partial_x u(\boldsymbol{x}))^\top\left(\dot{\boldsymbol{\nu}}_t + \boldsymbol{S}_t^\top \boldsymbol{\Xi}_t^{-1}(\boldsymbol{x} - \boldsymbol{\nu}_t)\right) + \frac{h}{2}\mathrm{tr}[(\partial^2_{xx}u(\boldsymbol{x}))(\dot{\boldsymbol{\Xi}}_t - (\boldsymbol{S}_t + \boldsymbol{S}_t^\top))] + ... \tag{102}$$

Subtracting $u(x)$ and taking the limit $h \to 0$, it is clear from the definition of the generator that the SDE drift is

$$\dot{\boldsymbol{\nu}}_t + \boldsymbol{S}_t^\top \boldsymbol{\Xi}_t^{-1}(\boldsymbol{x} - \boldsymbol{\nu}_t). \tag{103}$$

This takes the same form as the equation found in [5], however our formulae for $\boldsymbol{S}_t$ allow us to apply it to any linear reference SDE, not necessarily ones with scalar drift. Additionally, we empirically verify that for asymmetric $\boldsymbol{A}$ the matrix $\boldsymbol{S}_t$ is asymmetric. This contrasts with the symmetric nature of the drift for the gradient-type setting.

Now we want to work out what each of these terms are in practice. Effectively we need to compute $\dot{\mathfrak{A}}_t$, $\dot{\mathfrak{B}}_t$ and $(\partial_{t'}\boldsymbol{\Omega}_{t,t'})(t)$:

$$\dot{\mathfrak{A}}_t = (\boldsymbol{A}e^{-(T-t)\boldsymbol{A}} - \dot{\boldsymbol{\Gamma}}_t)\boldsymbol{\Sigma}_T^{1/2}, \tag{104}$$

$$\dot{\mathfrak{B}}_t = \dot{\boldsymbol{\Gamma}}_t\boldsymbol{\Sigma}_T^{1/2}, \tag{105}$$

$$\dot{\boldsymbol{\Gamma}}_t = \left(e^{t\boldsymbol{A}}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top e^{T\boldsymbol{A}^\top} - \boldsymbol{\Phi}_t\boldsymbol{A}^\top e^{(T-t)\boldsymbol{A}^\top}\right)\boldsymbol{\Phi}_T^{-1}, \tag{106}$$

$$(\partial_{t'}\boldsymbol{\Omega}_{t,t'})(t) = \boldsymbol{\Phi}_t\left[\boldsymbol{A}^\top - e^{(T-t)\boldsymbol{A}^\top}\boldsymbol{\Phi}_T^{-1}e^{(T-t)\boldsymbol{A}}\left(-\boldsymbol{A}\boldsymbol{\Phi}_t + e^{t\boldsymbol{A}}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top e^{t\boldsymbol{A}^\top}\right)\right]. \tag{107}$$

Remark on the case of time-varying coefficients In this case we can still assume without loss of generality that we still work on the time interval $[0, T]$, by shifting the time coordinate if necessary. Let $\boldsymbol{\Psi}(t, s)$ be the state transition matrix associated with $\boldsymbol{A}_t$ and $\boldsymbol{\Phi}_t$ be the solution at time t to (53). Then we still have $\boldsymbol{\Sigma}_T = \boldsymbol{\Phi}_T$ for the transition kernel covariance in the cost, i.e. the covariance started from a point mass. For the mean of the reference process started from $\boldsymbol{x}_0$, we have the generalised formula

$$\boldsymbol{\mu}_t(\boldsymbol{x}_0) = \boldsymbol{\Psi}(t, 0)\boldsymbol{x}_0 - \int_0^t \boldsymbol{\Psi}(t, s)\boldsymbol{A}_s\boldsymbol{m}_s\,\mathrm{d}s, \tag{108}$$

and we remark that its inverse $\boldsymbol{\mu}_t^{-1}$ always exists since $\boldsymbol{\Psi}$ is never singular:

$$\boldsymbol{\mu}_t^{-1}(\boldsymbol{y}) = \boldsymbol{\Psi}(t, 0)^{-1}\left(\boldsymbol{y} + \int_0^t \boldsymbol{\Psi}(t, s)\boldsymbol{A}_s\boldsymbol{m}_s\,\mathrm{d}s\right). \tag{109}$$

Then

$$\overline{\boldsymbol{a}} = \boldsymbol{\Sigma}_T^{-1/2}\boldsymbol{\mu}_T^{\boldsymbol{a}}, \tag{110}$$

$$\overline{\boldsymbol{\mathcal{A}}} = \boldsymbol{\Sigma}_T^{-1/2}\boldsymbol{\Psi}(t, 0)\boldsymbol{\mathcal{A}}\boldsymbol{\Psi}(t, 0)^\top\boldsymbol{\Sigma}_T^{-1/2}, \tag{111}$$

$$\overline{\boldsymbol{b}} = \boldsymbol{\Sigma}_T^{-1/2}\boldsymbol{b}, \tag{112}$$

$$\overline{\boldsymbol{\mathcal{B}}} = \boldsymbol{\Sigma}_T^{-1/2}\boldsymbol{\mathcal{B}}\boldsymbol{\Sigma}_T^{-1/2}. \tag{113}$$

Recall that the general expression for the $\mathbb{Q}$-bridge conditioned on $(0, \boldsymbol{x}_0), (T, \boldsymbol{x}_T)$ is

$$\boldsymbol{\mu}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_1)} = \boldsymbol{\mu}_t(\boldsymbol{x}_0) + \boldsymbol{\Phi}_t\boldsymbol{\Psi}(T, t)^\top\boldsymbol{\Phi}_T^{-1}(\boldsymbol{x}_T - \boldsymbol{\mu}_T(\boldsymbol{x}_0)). \tag{114}$$

Letting $\boldsymbol{x}_0 = \boldsymbol{\mu}_T^{-1}(\boldsymbol{\Sigma}_T^{1/2}\overline{\boldsymbol{x}}_0)$ and $\boldsymbol{x}_T = \boldsymbol{\Sigma}_T^{1/2}\overline{\boldsymbol{x}}_T$ and substituting the expression for $\boldsymbol{\mu}_t^{-1}$ we have

$$\begin{aligned}\boldsymbol{\mu}_{t|(T_0^{-1}(\overline{\boldsymbol{x}}_0),T_T^{-1}(\overline{\boldsymbol{x}}_T))} &= \left(\boldsymbol{\Psi}(t,T) - \boldsymbol{\Phi}_t\boldsymbol{\Psi}(T,t)^\top\boldsymbol{\Phi}_T^{-1}\right)\boldsymbol{\Sigma}_T^{1/2}\overline{\boldsymbol{x}}_0 \\ &+ \boldsymbol{\Phi}_t\boldsymbol{\Psi}(T,t)^\top\boldsymbol{\Phi}_T^{-1}\boldsymbol{\Sigma}_T^{1/2}\overline{\boldsymbol{x}}_T \\ &+ \int_t^T \boldsymbol{\Psi}(t,s)\boldsymbol{A}_s\boldsymbol{m}_s\mathrm{d}s.\end{aligned} \tag{115}$$

Set $\boldsymbol{\Gamma}_t = \boldsymbol{\Phi}_t\boldsymbol{\Psi}(T,t)^\top\boldsymbol{\Phi}_T^{-1}$, then

$$\mathfrak{A}_t = \left(\boldsymbol{\Psi}(t,T) - \boldsymbol{\Gamma}_t\right)\boldsymbol{\Sigma}_T^{1/2}, \tag{116}$$

$$\mathfrak{B}_t = \boldsymbol{\Gamma}_t\boldsymbol{\Sigma}_T^{1/2}, \tag{117}$$

$$\mathfrak{c}_t = \int_t^T \boldsymbol{\Psi}(t,s)\boldsymbol{A}_s\boldsymbol{m}_s\mathrm{d}s. \tag{118}$$

Case: Brownian motion Let's check this against the results of Bunne and Hsieh [5], who consider a class of reference processes

$$\mathrm{d}\boldsymbol{Y}_t = (c_t\boldsymbol{Y}_t + \alpha_t)\,\mathrm{d}t + g_t\,\mathrm{d}\boldsymbol{B}_t.$$

This corresponds to a special case of ours, where the drift is scalar-valued. Setting $g_t = \omega, \alpha_t = c_t = 0$, their results give us the marginal parameters of the GSB connecting $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$. In what follows, using their results and notations of [5, Table 1], we have that $r_t = t, \overline{r}_t = 1-t, \zeta = 0, \kappa(t,t') = \omega^2 t, \rho_t = t$. Then:

$$\boldsymbol{\mu}_t = \overline{r}_t\boldsymbol{\mu}_0 + r_t\boldsymbol{\mu}_1 + \zeta(t) - r_t\zeta(1), \tag{119}$$

$$= (1-t)\boldsymbol{\mu}_0 + t\boldsymbol{\mu}_1, \tag{120}$$

$$\boldsymbol{\Sigma}_t = \overline{r}_t^2\boldsymbol{\Sigma}_0 + r_t^2\boldsymbol{\Sigma}_1 + r_t\overline{r}_t(\boldsymbol{C}_\sigma + \boldsymbol{C}_\sigma^\top) + \kappa(t,t')(1-\rho_t)\mathbf{I} \tag{121}$$

$$= (1-t)^2\boldsymbol{\Sigma}_0 + t^2\boldsymbol{\Sigma}_1 + t(1-t)(\boldsymbol{C}_\sigma + \boldsymbol{C}_\sigma^\top) + \omega^2 t(1-t)\mathbf{I}. \tag{122}$$

Let us compute the same quantities using our formulae. We set $\boldsymbol{A} = 0, \boldsymbol{m} = 0$ and $T = 1$. Then, $\boldsymbol{\Phi}_t = t\omega^2\mathbf{I}, \boldsymbol{\Gamma}_t = t\mathbf{I}, \boldsymbol{\Sigma}_T^{1/2} = \omega\mathbf{I}$. Then,

$$\overline{\boldsymbol{\mathcal{A}}} = \omega^{-2}\boldsymbol{\mathcal{A}}, \qquad \overline{\boldsymbol{\mathcal{B}}} = \omega^{-2}\boldsymbol{\mathcal{B}}.$$

Also from the underbraced expressions, we have

$$\mathfrak{A}_t = \omega(1-t)\mathbf{I}, \qquad \mathfrak{B}_t = \omega t\mathbf{I}.$$

Assembling our results, we get

$$\begin{aligned}\mathbb{E}[\boldsymbol{X}_t] &= (\mathbf{I} - \boldsymbol{\Gamma}_t)\boldsymbol{a} + \boldsymbol{\Gamma}_t\boldsymbol{b} \\ &= (1-t)\boldsymbol{a} + t\boldsymbol{b} \\ \mathbb{V}[\boldsymbol{X}_t] &= \boldsymbol{\Omega}_t + \mathfrak{A}_t\overline{\boldsymbol{\mathcal{A}}}\mathfrak{A}_t^\top + \mathfrak{B}_t\overline{\boldsymbol{\mathcal{B}}}\mathfrak{B}_t^\top + \mathfrak{A}_t\overline{\boldsymbol{\mathcal{C}}}\mathfrak{B}_t^\top + \mathfrak{B}_t\overline{\boldsymbol{\mathcal{C}}}\mathfrak{A}_t^\top \\ &= \boldsymbol{\Omega}_t + (1-t)^2\boldsymbol{\mathcal{A}} + t^2\boldsymbol{\mathcal{B}} + \omega^2 t(1-t)(\overline{\boldsymbol{\mathcal{C}}} + \overline{\boldsymbol{\mathcal{C}}}^\top).\end{aligned}$$

Now $\overline{\boldsymbol{\mathcal{C}}}$ is the transport plan obtained from the *scaled* input Gaussians, of variance $\overline{\boldsymbol{\mathcal{A}}} = \boldsymbol{\mathcal{A}}/\omega^2, \overline{\boldsymbol{\mathcal{B}}} = \boldsymbol{\mathcal{B}}/\omega^2$ (in the case of Brownian reference process, there is no linear mapping via the flow map). For the scaled measures, EOT is calculated with $\varepsilon = \sigma^2 = 1$. It stands to reason that once we scale everything back, we should get $\overline{\boldsymbol{\mathcal{C}}} = \boldsymbol{\mathcal{C}}/\omega^2$. To verify this, recall that for the untransformed problem

$$\arg\min_{\pi\in\Pi(\alpha,\beta)} \frac{1}{2}\int \frac{1}{\omega^2}\|x-y\|^2\mathrm{d}\pi + \mathrm{H}(\pi|\alpha\otimes\beta) = \arg\min_{\pi\in\Pi(\alpha,\beta)} \frac{1}{2}\int \|x-y\|^2\mathrm{d}\pi + \omega^2\,\mathrm{H}(\pi|\alpha\otimes\beta).$$

Then for the RHS problem, we have that

$$\begin{aligned}\boldsymbol{\mathcal{C}} &= \boldsymbol{\mathcal{A}}^{1/2}(\boldsymbol{\mathcal{A}}^{1/2}\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{A}}^{1/2} - \frac{\omega^4}{4}\mathbf{I})^{1/2}\boldsymbol{\mathcal{A}}^{-1/2} - \frac{\omega^2}{2}\mathbf{I} \\ &= \omega^2(\overline{\boldsymbol{\mathcal{A}}}^{1/2}(\overline{\boldsymbol{\mathcal{A}}}^{1/2}\overline{\boldsymbol{\mathcal{B}}}\,\overline{\boldsymbol{\mathcal{A}}}^{1/2} + \frac{1}{4}\mathbf{I})^{1/2}\overline{\boldsymbol{\mathcal{A}}}^{-1/2} - \frac{1}{2}\mathbf{I}) \\ &= \omega^2\overline{\boldsymbol{\mathcal{C}}}.\end{aligned}$$

Also it is easy to verify that $\mathbf{\Omega}_t = \mathbf{\Sigma}_{t|(x_0,x_1)} = \omega^2 t(1-t)\mathbf{I}$. So,

$$\mathbb{V}[\boldsymbol{X}_t] = \omega^2 t(1-t)\mathbf{I} + (1-t)^2\boldsymbol{\mathcal{A}} + t^2\boldsymbol{\mathcal{B}} + t(1-t)(\boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{C}}^\top), \tag{123}$$

where $\boldsymbol{\mathcal{C}}$ is the EOT plan between the unscaled measures $\mathcal{N}(\boldsymbol{a}, \boldsymbol{\mathcal{A}}), \mathcal{N}(\boldsymbol{b}, \boldsymbol{\mathcal{B}})$ with $\varepsilon = \omega^2$. This is exactly the same result as the one derived in [5].

Case: Centered scalar OU process $\boldsymbol{A} = -\lambda\mathbf{I}, \boldsymbol{m} = \mathbf{0}$ From [5, Table 1], we have that $\overline{r}_t = \sinh(\lambda t)/\sinh(\lambda)$, $\overline{r}_t = \sinh(\lambda t)\coth(\lambda t) - \sinh(\lambda t)\coth(\lambda)$, $\zeta = 0$. The mean is then

$$\boldsymbol{\mu}_t = \overline{r}_t\boldsymbol{\mu}_0 + r_t\boldsymbol{\mu}_1. \tag{124}$$

With $\kappa(t,t') = \omega^2 e^{-\lambda t}\sinh(\lambda t)/\lambda$ and $\rho_t = e^{-\lambda(1-t)}\sinh(\lambda t)/\sinh(\lambda)$, the variance is

$$\mathbf{\Sigma}_t = \overline{r}_t^2\mathbf{\Sigma}_0 + r_t\mathbf{\Sigma}_1 + r_t\overline{r}_t(\boldsymbol{C}_{\sigma^\star} + \boldsymbol{C}_{\sigma^\star}^\top) + \kappa(t,t)(1-\rho_t)\mathbf{I}, \tag{125}$$

where $\sigma^{\star 2} = \omega^2\sinh(\lambda)/\lambda$. For convenience, we will check things term by term and using Mathematica.

Check first the mean. We have:

$$\mathbf{\Phi}_t = \omega^2\left(\frac{1-e^{-2\lambda t}}{2\lambda}\right)\mathbf{I}, \qquad \mathbf{\Gamma}_t = e^{-(1-t)\lambda}\left(\frac{1-e^{-2\lambda t}}{1-e^{-2\lambda}}\right)\mathbf{I} \tag{126}$$

Then, plugging into our formula,

$$\mathbb{E}[\boldsymbol{X}_t] = \underbrace{e^{-\lambda}(e^{(1-t)\lambda}\mathbf{I} - \mathbf{\Gamma}_t)}_{=\overline{r}_t}\boldsymbol{a} + \underbrace{\mathbf{\Gamma}_t}_{=r_t}\boldsymbol{b}. \tag{127}$$

Now the variance. We have that

$$\mathbf{\Sigma}_1 = \omega^2\left(\frac{e^{-\lambda}\sinh(\lambda)}{\lambda}\right)\mathbf{I}, \tag{128}$$

$$\overline{\boldsymbol{\mathcal{A}}} = \frac{\lambda e^{-\lambda}}{\omega^2\sinh(\lambda)}\boldsymbol{\mathcal{A}} = \alpha\boldsymbol{\mathcal{A}}, \tag{129}$$

$$\overline{\boldsymbol{\mathcal{B}}} = \frac{\lambda}{\omega^2 e^{-\lambda}\sinh(\lambda)}\boldsymbol{\mathcal{B}} = \beta\boldsymbol{\mathcal{B}}, \tag{130}$$

$$\mathfrak{A}_t = \left(e^{(1-t)\lambda}\mathbf{I} - \mathbf{\Gamma}_t\right)\mathbf{\Sigma}_1^{1/2}, \tag{131}$$

$$\mathfrak{B}_t = \mathbf{\Gamma}_t\mathbf{\Sigma}_1^{1/2}. \tag{132}$$

From here it is straightforward to verify that

$$\mathfrak{A}_t\overline{\boldsymbol{\mathcal{A}}}\mathfrak{A}_t^\top = \overline{r}_t^2\boldsymbol{\mathcal{A}}, \quad \mathfrak{B}_t\overline{\boldsymbol{\mathcal{B}}}\mathfrak{B}_t^\top = r_t^2\boldsymbol{\mathcal{B}}. \tag{133}$$

Now note that

$$\overline{\boldsymbol{\mathcal{C}}} = \overline{\boldsymbol{\mathcal{A}}}^{1/2}\left(\overline{\boldsymbol{\mathcal{A}}}^{1/2}\overline{\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{A}}}^{1/2} + \frac{1}{4}\mathbf{I}\right)^{1/2}\overline{\boldsymbol{\mathcal{A}}}^{-1/2} - \frac{1}{2}\mathbf{I} \tag{134}$$

$$= \sqrt{\alpha\beta}\left[\boldsymbol{\mathcal{A}}^{1/2}\left(\boldsymbol{\mathcal{A}}^{1/2}\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{A}}^{1/2} + \frac{(\alpha\beta)^{-1}}{4}\mathbf{I}\right)^{1/2}\boldsymbol{\mathcal{A}}^{-1/2} - \frac{(\alpha\beta)^{-1/2}}{2}\mathbf{I}\right] \tag{135}$$

$$= \sqrt{\alpha\beta}\boldsymbol{\mathcal{C}}_{1/\sqrt{\alpha\beta}}, \tag{136}$$

where $\boldsymbol{\mathcal{C}}_{1/\sqrt{\alpha\beta}}$ denotes the EOT covariance for entropic regularisation level $1/\sqrt{\alpha\beta}$. Then:

$$\mathfrak{A}_t\overline{\boldsymbol{\mathcal{C}}}\mathfrak{B}_t^\top + \mathfrak{B}_t\overline{\boldsymbol{\mathcal{C}}}^\top\mathfrak{A}_t^\top = \underbrace{(e^{\lambda(1-t)}\mathbf{I} - \mathbf{\Gamma}_t)\mathbf{\Gamma}_t\mathbf{\Sigma}_1\sqrt{\alpha\beta}}_{r_t\overline{r}_t}(\boldsymbol{\mathcal{C}}_{1/\sqrt{\alpha\beta}} + \boldsymbol{\mathcal{C}}_{1/\sqrt{\alpha\beta}}^\top) \tag{137}$$

It is also easy to check that

$$\frac{1}{\sqrt{\alpha\beta}} = \frac{\omega^2\sinh(\lambda)}{\lambda}.$$

Finally,

$$\mathbf{\Omega}_t = \mathbf{\Sigma}_{t|(x_0,x_1)} = \mathbf{\Phi}_t(\mathbf{I} - e^{-2\lambda(1-t)}\mathbf{\Phi}_t\mathbf{\Phi}_1^{-1}) = \kappa(t,t)(1-\rho_t)\mathbf{I}. \tag{138}$$

We have verified that all the terms in the expression for the variance agree.

A.5 Proof of Theorem 4

Proof. We follow the arguments of [48] with application to the generalised Schrödinger bridge – not much changes for $\mathbb{Q}$ as the reference and we reproduce in detail the arguments as follows. For the equality of gradients, it suffices to show this for the flow matching component of the loss, since the score matching component can be handled using the exact same arguments [48]. Let $\boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}$ be the conditional flow between $(\boldsymbol{x}_0,\boldsymbol{x}_T)$ and $\boldsymbol{u}_t$ be the SB flow defined by

$$\boldsymbol{u}_t(\boldsymbol{x}) = \mathbb{E}_{(\boldsymbol{x}_0,\boldsymbol{x}_T)\sim\pi}\left[\frac{p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})}{p_t(\boldsymbol{x})}\boldsymbol{u}_{t|(x_0,x_T)}(\boldsymbol{x})\right], \quad p_t(\boldsymbol{x}) = \mathbb{E}_{(\boldsymbol{x}_0,\boldsymbol{x}_T)\sim\pi}\, p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x}).$$

Let $\boldsymbol{u}_t^\theta$ be the neural flow approximation with parameter θ. Assuming that $p_t(\boldsymbol{x}) > 0$ for all $(t,\boldsymbol{x})$ we then have:

$$\nabla_\theta \mathbb{E}_{(\boldsymbol{x}_0,\boldsymbol{x}_T)\sim\pi}\mathbb{E}_{\boldsymbol{x}\sim p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}}\|\boldsymbol{u}_t^\theta(\boldsymbol{x}) - \boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\|^2 - \nabla_\theta\mathbb{E}_{\boldsymbol{x}\sim p_t(\boldsymbol{x})}\|\boldsymbol{u}_t^\theta(\boldsymbol{x}) - \boldsymbol{u}_t(\boldsymbol{x})\|^2 \tag{139}$$

$$= \nabla_\theta \mathbb{E}_{(\boldsymbol{x}_0,\boldsymbol{x}_T)\sim\pi}\mathbb{E}_{\boldsymbol{x}\sim p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}}\left[\|\boldsymbol{u}_t^\theta(\boldsymbol{x})\|^2 + \|\boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\|^2 - 2\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\rangle\right] \tag{140}$$

$$- \nabla_\theta\mathbb{E}_{\boldsymbol{x}\sim p_t(\boldsymbol{x})}\left[\|\boldsymbol{u}_t^\theta(\boldsymbol{x})\|^2 + \|\boldsymbol{u}_t(\boldsymbol{x})\|^2 - 2\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_t(\boldsymbol{x})\rangle\right] \tag{141}$$

$$= 2\nabla_\theta \mathbb{E}_{(\boldsymbol{x}_0,\boldsymbol{x}_T)\sim\pi}\mathbb{E}_{\boldsymbol{x}\sim p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}}\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_t(\boldsymbol{x})\rangle - \langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\rangle, \tag{142}$$

where in the last line terms not depending on θ are zero and we use the fact that $\mathbb{E}_{(\boldsymbol{x}_0,\boldsymbol{x}_T)}\mathbb{E}_{\boldsymbol{x}|(\boldsymbol{x}_0,\boldsymbol{x}_T)} f(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{x}} f(\boldsymbol{x})$. Now, using the relation between $\boldsymbol{u}_t$ and $\boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}$,

$$\mathbb{E}_{\boldsymbol{x}\sim p_t(\boldsymbol{x})}\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_t(\boldsymbol{x})\rangle = \int \mathrm{d}p_t(\boldsymbol{x})\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_t(\boldsymbol{x})\rangle \tag{143}$$

$$= \int \mathrm{d}p_t(\boldsymbol{x})\left\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \int \mathrm{d}\pi(\boldsymbol{x}_0,\boldsymbol{x}_T)\frac{p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})}{p_t(\boldsymbol{x})}\boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\right\rangle \tag{144}$$

$$= \int \mathrm{d}\pi(\boldsymbol{x}_0,\boldsymbol{x}_T)\int \mathrm{d}\boldsymbol{x}\, p_t(\boldsymbol{x})\left\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \frac{p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})}{p_t(\boldsymbol{x})}\boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\right\rangle \tag{145}$$

$$= \int \mathrm{d}\pi(\boldsymbol{x}_0,\boldsymbol{x}_T)\int \mathrm{d}p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\ \langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\rangle \tag{146}$$

$$= \mathbb{E}_{(\boldsymbol{x}_0,\boldsymbol{x}_T)\sim\pi}\mathbb{E}_{\boldsymbol{x}\sim p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}}\langle \boldsymbol{u}_t^\theta(\boldsymbol{x}), \boldsymbol{u}_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})\rangle. \tag{147}$$

So we conclude that the two gradients are equal. Clearly, the *unconditional* loss can be rewritten over candidate flow and score fields $(\hat{\boldsymbol{u}},\hat{\boldsymbol{s}})$ as

$$(\hat{\boldsymbol{u}},\hat{\boldsymbol{s}}) \mapsto \|\hat{\boldsymbol{u}} - \boldsymbol{u}\|^2_{L^2(\mathrm{d}p_t(\boldsymbol{x})\mathrm{d}t)} + \|\lambda_t(\hat{\boldsymbol{s}} - \boldsymbol{s})\|^2_{L^2(\mathrm{d}p_t(\boldsymbol{x})\mathrm{d}t)},$$

where we denote by $\|h_t(\boldsymbol{x})\|^2_{L^2(\mathrm{d}p_t(\boldsymbol{x})\mathrm{d}t)} = \int_0^1 \mathrm{d}t \int \mathrm{d}p_t(\boldsymbol{x})\|h_t(\boldsymbol{x})\|^2_{L^2(\mathbb{R}^d)}$ for a test function $h_t : [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$. This loss is zero iff $\hat{\boldsymbol{u}} = \boldsymbol{u}$ and $\hat{\boldsymbol{s}} = \boldsymbol{s}$ $(\mathrm{d}p_t \times \mathrm{d}t)$-almost everywhere.

Let $\mathbb{P}$ denote the law of the Schrödinger bridge as per (SBP-dyn) and write $p_t(\boldsymbol{x}) = \mathbb{P}_t$ to mean its marginal at time t. Then in (2) and for $\mathbb{Q}$ a mvOU process (1), identifying:

- $\mathrm{d}\mathbb{P}^\star_{0T} = \pi$ where π is prescribed in Proposition 1, and
- $\mathbb{Q}_t^{x_0,x_T} = p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}(\boldsymbol{x})$ where $p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}$ is defined as in Theorem 2,

substituting all these into (2) one has

$$p_t(\boldsymbol{x}) = \int \mathrm{d}\pi(\boldsymbol{x}_0,\boldsymbol{x}_T)p_{t|(\boldsymbol{x}_0,\boldsymbol{x}_T)}.$$

Consequently, there exists $\boldsymbol{u}_t(\boldsymbol{x})$ such that $\partial_t p_t(\boldsymbol{x}) = -\nabla\cdot(p_t(\boldsymbol{x})u_t(\boldsymbol{x}))$. Writing $\boldsymbol{s}_t(\boldsymbol{x}) = \nabla_x \log p_t(\boldsymbol{x})$ to be the score, recognising terms from the probability flow ODE, it follows that that the SDE

$$\mathrm{d}\boldsymbol{X}_t = (\boldsymbol{u}_t(\boldsymbol{X}_t) + \boldsymbol{D}\boldsymbol{s}_t(\boldsymbol{X}_t))\mathrm{d}t + \boldsymbol{\sigma}\mathrm{d}\boldsymbol{B}_t \tag{148}$$

generates the marginals $p_t(\boldsymbol{x})$ of the Schrödinger bridge. Since $\mathbb{P}$ is characterised as a mixture of $\mathbb{Q}$-bridges, it follows that $\boldsymbol{X}_t$ defined by the SDE (148) generates the Markovization of $\mathbb{P}$ (see Appendix B of [48]). Moreover, $\mathbb{P}$ is the *unique* process that is both Markov and a mixture of $\mathbb{Q}$-bridges [26, 42]. □

A.6 Additional background

Hamilton-Jacobi-Bellman equation Consider the *deterministic* control problem on $[0, T]$:

$$V_0(x_0) = \min_u \int_0^T C(x_t, u_t)\,\mathrm{d}t + D(x_T). \tag{149}$$

subject to a dynamics $\dot{x}_t = F(x_t, u_t)$. $V_t(x)$ is the *value* function for the problem starting at (t, x) up to the final time T. The function $x_T \mapsto D(x_T)$ specifies a cost on the final value. The corresponding Hamilton-Jacobi-Bellman (HJB) equation [22, Section 3.11] is

$$\partial_t V_t(x) + \min_{u_t} \{\langle \nabla_x V_t(x), F(x, u_t(x))\rangle + C(x, u_t(x))\} = 0, \tag{150}$$

subject to the final boundary condition $V_T(x) = D(x)$. A heuristic derivation is as follows. Considering a time interval $(t, t+\delta t)$ and a path x_t, it's clear that

$$V_t(x_t) = \min_u \left\{ \int_t^{t+\delta t} C(x_s, u_s)\mathrm{d}s + V_{t+\delta t}(x_{t+\delta t}) \right\}. \tag{151}$$

Using Taylor expansion and the constraint $\dot{x}_t = F(x_t, u_t)$ we have to leading order that

$$V_{t+\delta t}(x_{t+\delta t}) = V_t(x_t) + (\partial_t V_t(x_t) + \langle \nabla_x V_t(x_t), F(x_t, u_t)\rangle)\,\delta t + O(\delta t^2). \tag{152}$$

Substituting back and also approximating the integral, we find that

$$V_t(x_t) = \min_u \ C(x_t, u_t)\delta t + V_t(x_t) + (\partial_t V_t(x_t) + \langle \nabla_x V_t(x_t), F(x_t, u_t)\rangle)\,\delta t \tag{153}$$

Cancelling terms, rearranging and taking a limit $\delta t \downarrow 0$, we get the desired result.

Stochastic Hamilton-Jacobi-Bellman Now we consider the *stochastic* variant of the HJB, in which case X_t is driven by an SDE of the form

$$\mathrm{d}X_t = F(X_t, u_t)\mathrm{d}t + \sigma_t \mathrm{d}B_t. \tag{154}$$

The value function is thus in expectation:

$$V_0(x_0) = \min_u \mathbb{E}\left\{ \int_0^T C(X_t, u_t)\mathrm{d}t + D(X_T) \right\}. \tag{155}$$

Carrying out the same expansion as before, we have

$$V_t(x_t) = \min_u \mathbb{E}\left\{ \int_t^{t+\delta t} C(X_s, u_s)\mathrm{d}s + V_{t+\delta t}(X_{t+\delta t}) \right\}. \tag{156}$$

Note that in the above, $(X_{t+\delta t}|X_t = x_t)$ is a random variable and hence so is $V_{t+\delta t}(X_{t+\delta t})$ which is why it appears in the expectation. Using Ito's formula, to expand this, we get

$$\begin{aligned} V_{t+\delta t}(X_{t+\delta t}) &= V_t(X_t) + \left[\partial_t V_t(X_t) + \frac{1}{2}\nabla \cdot (\sigma_t \sigma_t^\top \nabla_x V_t(X_t))\right]\mathrm{d}t + \langle \nabla_x V_t(X_t), \mathrm{d}X_t\rangle \\ &= V_t(X_t) + \left[\partial_t V_t(X_t) + \frac{1}{2}\nabla \cdot (\sigma_t \sigma_t^\top \nabla_x V_t(X_t)) + \langle \nabla_x V_t(X_t), F(X_t, u_t)\rangle\right]\mathrm{d}t + \langle \nabla_x V_t(X_t), \sigma_t \mathrm{d}B_t\rangle. \end{aligned}$$

Plugging this in, cancelling terms, and noting that the final term has zero expectation, we find that

$$0 = \partial_t V_t(X_t) + \min_u \{\mathcal{A}[V_t](X_t, u_t) + C(X_t, u_t)\} \tag{157}$$

where $\mathcal{A}$ is the generator of the SDE governing X_t, i.e.

$$\mathcal{A}[f](x, u) = \langle F(x, u), \nabla_x f(x)\rangle + \frac{1}{2}\nabla \cdot (\sigma_t \sigma_t^\top \nabla_x f). \tag{158}$$

The HJB equation for stochastic control problem is therefore

$$0 = \partial_t V_t(x) + \min_u \{\mathcal{A}[V_t](x, u_t(x)) + C(x, u_t(x))\}, \tag{159}$$

i.e. this is the same as the HJB for the deterministic case, except with a diffusive term arising from the stochasticity.

B Experiment details

For (mvOU, BM)-OTFM and IPFP, all computations were carried out by CPU (8x Intel Xeon Gold 6254). NLSB and SBIRR computations were accelerated using a single NVIDIA L40S GPU.

B.1 Gaussian benchmarking

We consider settings of dimension $d \in \{2, 5, 10, 25, 50\}$ with data consisting of an initial and terminal Gaussian distribution and a mvOU reference process constructed as follows. We sample a $d \times 2$ submatrix $\boldsymbol{U}$ from a random $d \times d$ orthogonal matrix. Then for the reference process, $\boldsymbol{A}$ is constructed as the $d \times d$ matrix

$$\boldsymbol{A} = \boldsymbol{U} \begin{bmatrix} 0 & 1 \\ -2.5 & 0 \end{bmatrix} \boldsymbol{U}^\top.$$

Similarly, we let $\boldsymbol{m} = \boldsymbol{U} \begin{bmatrix} 1 & -1 \end{bmatrix}^\top$. For the marginals, we take $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ where

$$\boldsymbol{\mu}_0 = \boldsymbol{U} \begin{bmatrix} -2.5 & -0.5 \end{bmatrix}^\top, \qquad \boldsymbol{\Sigma}_0 = \boldsymbol{U} \begin{bmatrix} 0.1 & 0.005 \\ 0.005 & 0.1 \end{bmatrix} \boldsymbol{U}^\top + 0.1\mathbf{I}.$$

$$\boldsymbol{\mu}_1 = \boldsymbol{U} \begin{bmatrix} 0.5 & 2.5 \end{bmatrix}^\top, \qquad \boldsymbol{\Sigma}_1 = \boldsymbol{U} \begin{bmatrix} 1.1 & -2 \\ -2 & 1.1 \end{bmatrix} \boldsymbol{U}^\top + 0.1\mathbf{I}.$$

For each d, the initial and final marginals are approximated by $N = 128$ samples $\{\boldsymbol{x}_0^i\}_{i=1}^N \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, $\{\boldsymbol{x}_1^i\}_{i=1}^N \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$. We fix $\boldsymbol{\sigma} = \boldsymbol{I}$ and compute the exact marginal parameters $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)_{t \in [0,1]}$ using the formulas of Theorem 3 and reference parameters $(\boldsymbol{A}, \boldsymbol{m})$. Additionally, we compute the SDE drift $\boldsymbol{v}_{\mathrm{SB}}(t, \boldsymbol{x})$ of the mvOU-GSB using (17).

(mvOU, BM)-OTFM We apply mvOU-OTFM as per Algorithm 1 to learn a conditional flow matching approximation to the Schrödinger bridge. We choose to parameterise the probability flow and score fields using two feed-forward neural networks $\boldsymbol{u}_\theta(t, \boldsymbol{x}) = \mathrm{NN}_\theta(d+1, d)(t, \boldsymbol{x})$ and $\boldsymbol{s}_\varphi(t, \boldsymbol{x}) = \mathrm{NN}_\varphi(d+1, d)(t, \boldsymbol{x})$, each with $[64, 64, 64]$ hidden dimensions and ReLU activations. We use a batch size of 64 and learning rate 10^{-2} for 2,500 iterations using the AdamW optimizer. For BM-OTFM, the same training procedure was used except the reference process was taken to be Brownian motion with unit diffusivity.

IPFP We use the IPFP implementation provided by the authors of [49], specifically using the variant of their algorithm based on Gaussian process approximations to the Schrödinger bridge drift [49, Algorithm 2]. We provide the mvOU reference process as the prior drift function, i.e. $\boldsymbol{x} \mapsto \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{m})$ and employ an exponential kernel for the Gaussian process vector field approximation. We run the IPFP algorithm for 10 iterations, which is double that used in the original publication. Since IPFP explicitly constructs forward and reverse processes, at each time $0 \leqslant t \leqslant 1$, IPFP outputs *two* estimates of the Schrödinger bridge marginal, one from each process. We consider both outputs and distinguish between them using the $(\rightarrow, \leftarrow)$ symbols in Table 1. For the vector field estimate we employ only the forward drift.

NLSB We use the NLSB implementation provided by the authors of [23]. Following [12, Section 4.6], for a mvOU reference process $\mathrm{d}\boldsymbol{X}_t = \boldsymbol{f}_t \,\mathrm{d}t + \sigma \,\mathrm{d}\boldsymbol{B}_t$, the generalised SB problem (SBP-dyn) can be rewritten as a stochastic control problem

$$\min_{\boldsymbol{u}_t} \mathbb{E}\left[\int_0^1 \frac{1}{2}\|\boldsymbol{u}_t\|_2^2 \,\mathrm{d}t\right], \text{ subject to } \mathrm{d}\boldsymbol{X}_t = (\boldsymbol{f}_t + \boldsymbol{u}_t)\,\mathrm{d}t + \sigma \,\mathrm{d}\boldsymbol{B}_t.$$

On the other hand, the Lagrangian SB problem [23, Definition 3.1] is

$$\min_{\boldsymbol{v}_t} \mathbb{E}\left[\int_0^1 L(t, \boldsymbol{x}, \boldsymbol{v}_t(\boldsymbol{x}))\mathrm{d}t\right], \text{ subject to } \mathrm{d}\boldsymbol{X}_t = \boldsymbol{v}_t \,\mathrm{d}t + \sigma \,\mathrm{d}\boldsymbol{B}_t.$$

This shows that the appropriate Lagrangian to use is $L(t, \boldsymbol{x}, \boldsymbol{v}_t(\boldsymbol{x})) = \frac{1}{2}\|\boldsymbol{v}_t(\boldsymbol{x}) - \boldsymbol{f}_t(\boldsymbol{x})\|_2^2$. We apply this in the mvOU setting by taking $\boldsymbol{f}_t(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{m})$. The NLSB approach backpropagate through solution of the SDE to directly learn a neural approximation of the SB drift $\boldsymbol{v}_t$. We train NLSB the same hyperparameters as used in the original publication, using the Adam optimizer with learning rate 10^{-3} for a total of 2,500 epochs.

Metrics We measure both the approximation error of Gaussian Schrödinger bridge marginals as well as the error in vector field estimation. For the marginal reconstruction, for each method and for each $0 \leqslant t \leqslant 1$ we sample points integrated forward in time and compute the mean $\hat{\boldsymbol{\mu}}_t$ and variance $\hat{\boldsymbol{\Sigma}}_t$ from samples. We then compare to the ground truth marginals $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ computed from exact formulas using the Bures-Wasserstein metric (62). For the vector field, for each method and for each time $0 \leqslant t \leqslant 1$ we sample $M = 1024$ points from the ground truth SB marginal $\boldsymbol{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ and empirically estimate $\|\hat{\boldsymbol{v}}_{\mathrm{SB}} - \boldsymbol{v}_{\mathrm{SB}}\|_{\mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)}$ from samples. All experiments were repeated over 5 independent runs and summary statistics for metrics are shown in Table 1.

B.2 Gaussian mixture example

We consider initial and terminal marginals sampled from Gaussian mixtures. At time $t = 0$, we sample from

$$\mathcal{N}\left(\boldsymbol{U}\begin{bmatrix}-0.5 & -0.5\end{bmatrix}^\top, 0.01\boldsymbol{U}\boldsymbol{U}^\top\right), \qquad \mathcal{N}\left(\boldsymbol{U}\begin{bmatrix}0.5 & 0.5\end{bmatrix}^\top, 0.0625\boldsymbol{U}\boldsymbol{U}^\top\right)$$

in a 1:1 ratio, and at time $t = 1$ we sample in the same fashion from

$$\mathcal{N}\left(\boldsymbol{U}\begin{bmatrix}-2.5 & -2.5\end{bmatrix}^\top, 0.01\boldsymbol{U}\boldsymbol{U}^\top\right), \qquad \mathcal{N}\left(\boldsymbol{U}\begin{bmatrix}2.5 & 2.5\end{bmatrix}^\top, 0.25\boldsymbol{U}\boldsymbol{U}^\top\right).$$

Here, the matrix $\boldsymbol{U}$ and reference process parameters $(\boldsymbol{A}, \boldsymbol{m})$ are the same as used for the previous Gaussian example for $d = 10$. All other training details are the same as for the Gaussian example.

B.3 Repressilator example

Based on the system studied in [41], we simulate stochastic trajectories from the system

$$\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \left(\frac{\beta}{1 + (x_3/k)^n} - \gamma x_1\right) \mathrm{d}t + \sigma \, \mathrm{d}B_t^{(1)} \\
\frac{\mathrm{d}x_2}{\mathrm{d}t} &= \left(\frac{\beta}{1 + (x_1/k)^n} - \gamma x_2\right) \mathrm{d}t + \sigma \, \mathrm{d}B_t^{(2)}, \qquad \begin{bmatrix}\beta \\ n \\ k \\ \gamma \\ \sigma\end{bmatrix} = \begin{bmatrix}10 \\ 3 \\ 1 \\ 1 \\ 0.1\end{bmatrix}, \\
\frac{\mathrm{d}x_3}{\mathrm{d}t} &= \left(\frac{\beta}{1 + (x_2/k)^n} - \gamma x_3\right) \mathrm{d}t + \sigma \, \mathrm{d}B_t^{(3)},
\end{aligned} \tag{160}$$

with initial condition $\mathcal{N}([1, 1, 2], 0.01\mathbf{I})$ and simulated for the time interval $t \in [0, 10]$ using the Euler-Maruyama discretisation. Snapshots were sampled at $T = 10$ time points evenly spaced on $[0, 10]$, each comprising of 100 samples.

We employ Algorithm 2 to this data, running for 5 iterations starting from an initial Brownian reference process $\boldsymbol{A} = \mathbf{0}, \boldsymbol{m} = \mathbf{0}$. For the inner loop running mvOU-OTFM (Algorithm 1), we parameterise the probability flow and score as in the Gaussian example, with feed-forward networks of hidden dimension $[64, 64, 64]$ and ReLU activations. We run mvOU-OTFM for 1,000 iterations using the AdamW optimiser, a batch size of 64 and a learning rate of 10^{-2}. At each step of the outer loop, we employ ridge regression to fit the updated mvOU reference parameters $(\boldsymbol{A}, \boldsymbol{m})$. We do this using the standard `RidgeCV` method implemented in the `scikit-learn` package, which automatically selects the regularisation parameter.

We also carry out hold-one-out runs where for $2 \leqslant i \leqslant 9$ (i.e. all time-points except for the very first and last), Algorithm 2 is applied to $T - 1$ snapshots with the snapshot at t_i held out. Once the mvOU reference parameters are learned, forward integration of the learned mvOU-SB is used to predict the marginal at t_i. We report the reconstruction error in terms of the earth-mover distance (EMD) and energy distance [38]. Table 2 shows results averaged over held-out timepoints, and full results (split by timepoint) are shown in Table 4.

Since the system marginals are unimodal we reason that they can be reasonably well approximated by Gaussians. For each $2 \leqslant i \leqslant 9$ we fit multivariate Gaussians to the snapshots at t_{i-1}, t_i, t_{i+1} and use the results of Theorem 3 with the fitted mvOU reference output by Algorithm 2 to solve the mvOU-GSB between $p_{t_{i-1}}, p_{t_{i+1}}$. This is illustrated in Figure 4 for $i = 4$, and we show the full results for all timepoints in Figure 6 in comparison with the standard Brownian GSB.

Algorithm 2 Iterated reference fitting with mvOU-OTFM

Input: Samples $\{\boldsymbol{x}_i^{t_j}\}_{i=1}^{N_j}$ from multiple snapshots at times $\{t_j\}_{j=1}^{T}$, initial mvOU reference parameters $(\boldsymbol{A}, \boldsymbol{m})$, diffusivity $\boldsymbol{D} = \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top$
Initialise: Probability flow field $\boldsymbol{u}_t^\theta(\boldsymbol{x})$, score field $\boldsymbol{s}_t^\varphi(\boldsymbol{x})$.
Define: $\texttt{FitReference}(\boldsymbol{X}, \boldsymbol{V}) := \arg\min_{\boldsymbol{A},\boldsymbol{m}} \|\boldsymbol{V} - \boldsymbol{A}(\boldsymbol{X} - \boldsymbol{m})\|_2^2 + \lambda\|\boldsymbol{A}\|_F^2 + \gamma\|\boldsymbol{m}\|^2$
$\hat{\rho}_j \leftarrow N^{-1}\sum_{i=1}^{N} \delta_{\boldsymbol{x}_i^{t_j}}, 1 \leqslant j \leqslant T$ *Form empirical marginals*
while not converged **do**
  $(\boldsymbol{u}_t^\theta, \boldsymbol{s}_t^\theta) \leftarrow \texttt{fitOTFM}_{(\boldsymbol{A},\boldsymbol{m},\boldsymbol{D})}(\hat{\rho}_1, \dots, \hat{\rho}_T)$ *Fit flow and score with reference parameters*
  $\boldsymbol{v}_i^{t_j} \leftarrow (\boldsymbol{u}_{t_j}^\theta + \boldsymbol{D}\boldsymbol{s}_{t_j}^\varphi)(\boldsymbol{x}_i^{t_j}), \quad 1 \leqslant i \leqslant N_j, 1 \leqslant j \leqslant T$ *Get SDE drift*
  $\boldsymbol{A}, \boldsymbol{m} \leftarrow \texttt{FitReference}(\{(\boldsymbol{v}_i^{t_j})_{i=1}^{N_i}\}_{j=1}^{T}, \{(\boldsymbol{x}_i^{t_j})_{i=1}^{N_i}\}_{j=1}^{T})$ *Update reference parameters*
end while

		Leave-one-out marginal interpolation error						
Error metric	t	Iterate 0	Iterate 1	Iterate 2	Iterate 3	Iterate 4	SBIRR (mvOU)	SBIRR (MLP)
EMD	1	3.59 ± 0.17	3.14 ± 0.12	2.28 ± 0.15	2.08 ± 0.11	$\mathbf{2.02 \pm 0.11}$	2.24 ± 0.28	2.71 ± 0.41
	2	5.20 ± 0.47	2.59 ± 0.29	1.62 ± 0.28	1.27 ± 0.25	$\mathbf{1.13 \pm 0.16}$	3.13 ± 0.62	2.29 ± 0.92
	3	3.23 ± 0.24	1.42 ± 0.18	1.10 ± 0.14	0.86 ± 0.08	$\mathbf{0.83 \pm 0.10}$	2.67 ± 0.85	1.39 ± 0.55
	4	1.48 ± 0.20	0.52 ± 0.05	0.47 ± 0.05	$\mathbf{0.47 \pm 0.06}$	0.48 ± 0.06	1.38 ± 0.46	0.94 ± 0.28
	5	2.50 ± 0.40	1.43 ± 0.65	$\mathbf{1.12 \pm 0.32}$	1.29 ± 0.11	1.21 ± 0.13	1.63 ± 0.17	1.40 ± 0.71
	6	6.18 ± 0.41	3.42 ± 0.50	2.18 ± 0.28	1.91 ± 0.38	$\mathbf{1.75 \pm 0.17}$	2.40 ± 0.32	1.96 ± 1.41
	7	2.56 ± 0.25	3.09 ± 1.53	2.13 ± 0.46	2.23 ± 0.55	1.93 ± 0.50	1.45 ± 0.20	$\mathbf{0.51 \pm 0.13}$
	8	2.29 ± 0.26	2.12 ± 0.09	1.82 ± 0.33	1.81 ± 0.31	$\mathbf{1.81 \pm 0.16}$	1.93 ± 0.61	2.14 ± 0.40
Energy	1	3.00 ± 0.09	2.75 ± 0.07	2.24 ± 0.10	2.10 ± 0.07	$\mathbf{2.05 \pm 0.08}$	2.20 ± 0.18	2.56 ± 0.26
	2	3.53 ± 0.21	2.27 ± 0.19	1.66 ± 0.20	1.38 ± 0.21	$\mathbf{1.26 \pm 0.13}$	2.61 ± 0.33	2.05 ± 0.61
	3	2.29 ± 0.17	1.31 ± 0.16	1.00 ± 0.10	0.80 ± 0.06	$\mathbf{0.78 \pm 0.07}$	2.10 ± 0.52	1.25 ± 0.43
	4	0.93 ± 0.14	0.25 ± 0.06	0.17 ± 0.04	0.15 ± 0.02	$\mathbf{0.14 \pm 0.01}$	0.99 ± 0.38	0.84 ± 0.25
	5	1.10 ± 0.18	0.53 ± 0.27	$\mathbf{0.45 \pm 0.13}$	0.55 ± 0.04	0.51 ± 0.07	0.64 ± 0.10	0.66 ± 0.37
	6	2.49 ± 0.12	1.22 ± 0.10	1.13 ± 0.15	1.01 ± 0.22	0.91 ± 0.11	1.43 ± 0.17	$\mathbf{0.90 \pm 0.64}$
	7	0.97 ± 0.13	1.39 ± 0.65	1.02 ± 0.21	1.05 ± 0.25	0.91 ± 0.25	0.65 ± 0.09	$\mathbf{0.17 \pm 0.08}$
	8	0.54 ± 0.11	0.57 ± 0.03	0.56 ± 0.11	0.55 ± 0.10	0.55 ± 0.04	0.48 ± 0.08	$\mathbf{0.36 \pm 0.06}$

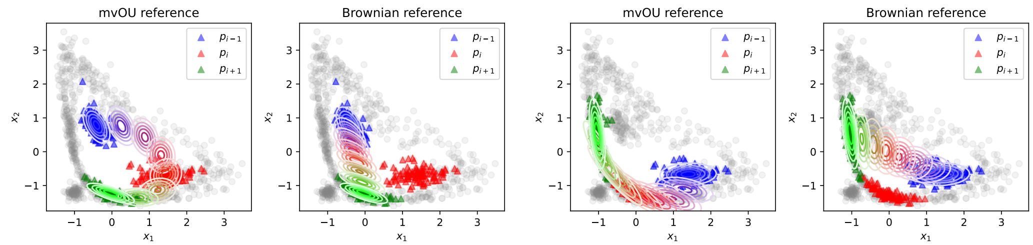
Table 4: Full results for repressilator example

SBIRR We apply SBIRR [41] using the implementation provided with the original publication, which notably includes an improved implementation of IPFP [49] that utilises GPU acceleration. We provide the reference vector field $\boldsymbol{x} \mapsto \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{m})$ and seek to learn a reference process in one of two families: (1) mvOU processes, i.e. we consider the family of reference drifts $\hat{\boldsymbol{A}}(\boldsymbol{x} - \hat{\boldsymbol{m}})$ where $\hat{\boldsymbol{A}}, \hat{\boldsymbol{m}}$ are to be fit, or (2) general drifts, i.e. we parameterise the drift using a feed-forward neural network with hidden dimensions $[64, 64, 64]$. For each choice of reference family, we run Algorithm 1 of [41] for 5 outer iterations and 10 inner IPFP iterations, as was also done in the original paper.

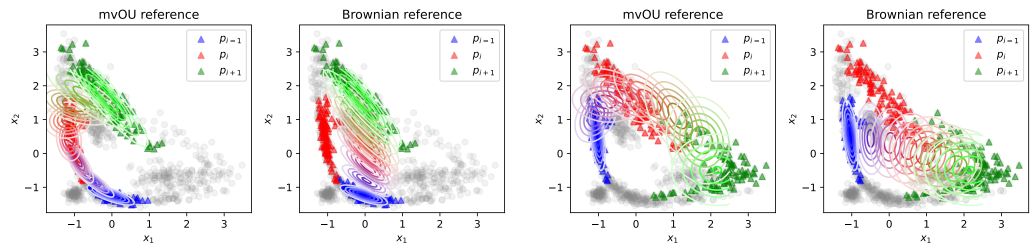
B.4 Cell cycle scRNA-seq

The metabolic labelled cell cycle dataset of [4] is obtained and preprocessed following the tutorial available with the Dynamo [37] package. This gives a dataset of $N = 2,793$ cells, embedded in 30 PCA dimensions. In addition to transcriptional state $\{\boldsymbol{x}_i\}_{i=1}^{N}$, Dynamo uses metabolic labelling data to predict the transcriptional velocity $\{\hat{\boldsymbol{v}}_i\}_{i=1}^{N}$ for each cell.

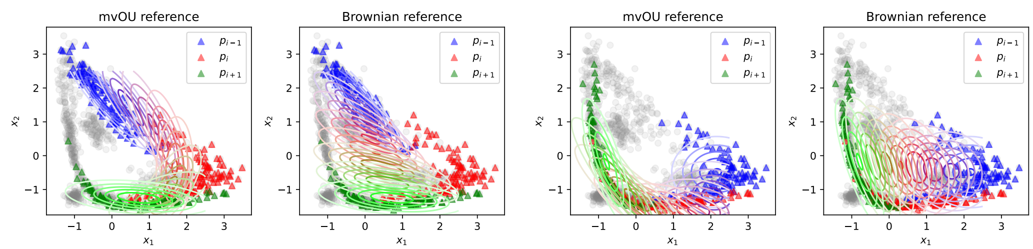
To fit the reference process parameters $(\boldsymbol{A}, \boldsymbol{m})$, we use again ridge regression via the `RidgeCV` method in `scikit-learn`. We train mvOU-OTFM using Algorithm 1 with $\sigma = 0.3$, parameterising the probability flow and score as previously using feed-forward networks of hidden dimensions $[64, 64, 64]$ and train with a batch size of 64, learning rate of 10^{-2} for a total of 1,000 iterations.



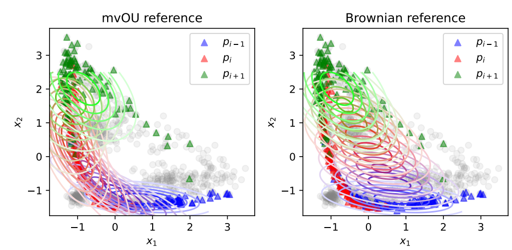
(a) $i = 1$ (b) $i = 2$



(c) $i = 3$ (d) $i = 4$



(e) $i = 5$ (f) $i = 6$



(g) $i = 7$

Figure 6: **Repressilator mvOU-GSB interpolation.** Using the learned mvOU reference process, we interpolate between p_{i-1} (blue) and p_{i+1} (green). Middle timepoint p_i is shown in red.

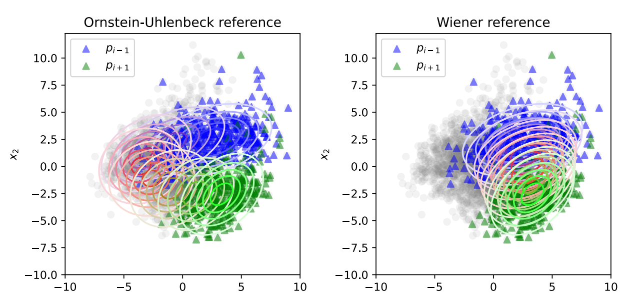


Figure 7: **Cell cycle mvOU-GSB interpolation.** Using the learned mvOU reference process and scale factor $\gamma = 50$, we interpolate between the first snapshot p_1 (blue) and last snapshot p_T (green). All computations are done in $d = 30$ and shown in leading 2 PCs.