

# Flow Matching on Gaussians: When is the Flow-Matching Map the Optimal-Transport Map?

Noogram Labs — cognition fleet  
flow-matching-gaussians

June 18, 2026

## Abstract

A stochastic interpolant  $X_t = a(t)X_0 + b(t)X_1$  between two centred Gaussians induces, through its conditional velocity field, a deterministic *flow-matching map*  $T_1$  pushing  $\mathcal{N}(0, \Sigma_0)$  onto  $\mathcal{N}(0, \Sigma_1)$ . The quadratic-cost optimal-transport (Bures–Wasserstein) problem between the same two Gaussians produces its own map  $T_{\text{OT}}$ . The two are easily conflated: they share domain, codomain, *and* pushforward, and on the favourite toy cases (isotropic covariances, scalar variances) they coincide. We show this agreement is an accident of commutation. For symmetric positive-definite endpoints the following are equivalent:  $[\Sigma_0, \Sigma_1] = 0$ ;  $T_1$  is symmetric;  $T_1 = T_{\text{OT}}$ ; and  $T_1 = \Sigma_1^{1/2}\Sigma_0^{-1/2}$ . Generically  $[\Sigma_0, \Sigma_1] \neq 0$ , the flow-matching map is non-symmetric and *not* the optimal-transport map, and the equality locus is a measure-zero subset of the cone of covariance pairs. Symmetry is the discriminant:  $T_1$  is the time-ordered solution of a linear ODE, while  $T_{\text{OT}}$  is the unique symmetric representative of the same family of pushforwards. We further prove that  $T_1$  is *schedule-independent* unconditionally — a flatness (zero-curvature) phenomenon, not a property of the commuting case alone. A closed-form three-Dirac mixture serves as the empirical illustration; covariance-ellipse experiments visualise the rotational “fingerprint” the non-commuting path leaves on  $T_1$ .

## 1 Introduction

*One sentence: the identification of flow matching with optimal transport is an accident of commuting toy cases.*

Generative modelling by *flow matching* and by *optimal transport* (OT) are often discussed as if they were two names for one object. Both move a simple source law to a target law; both, in the simplest examples one draws on a whiteboard, produce the same map. Stochastic interpolants [ABV23; AV23] build a path  $X_t = a(t)X_0 + b(t)X_1$  between samples of the two laws and learn the marginal velocity field  $v(x, t) = \mathbb{E}[\dot{a}X_0 + \dot{b}X_1 \mid X_t = x]$ ; integrating  $\dot{x} = v(x, t)$  transports the source to the target. This is the construction underlying flow matching [Lip+23; Lip+24], rectified flow [LGL23], and the probability-flow ODE of score-based diffusion [Son+21]. Optimal transport, by contrast, picks out the *cost-minimising* coupling; for the quadratic cost the Monge map is the gradient of a convex function [Bre91], and between Gaussians it is the closed-form symmetric positive-definite (SPD) Bures–Wasserstein map [BJL19; OP82; Tak11].

The two constructions agree on isotropic Gaussians and on scalar variances — exactly the cases on which intuition is built. We show that this is the exception, not the rule. For the independent-coupling Gaussian interpolant the flow-matching map equals the OT map *if and only if* the endpoint covariances commute, a measure-zero condition. The mechanism is sharp and geometric:  $T_1$  is the *time-ordered* exponential of the (matrix-valued) velocity generator, whereas  $T_{\text{OT}}$  is the unique *symmetric* pushforward between the two Gaussians. When the covariances fail to commute the time-ordering survives and  $T_1$  acquires a non-trivial rotational part — the path leaves a fingerprint on the endpoint map. Symmetry of  $T_1$  is precisely the discriminant.

## Contributions.

1. A single main theorem ([Theorem 1](#)): for SPD endpoints, the four statements “commute /  $T_1$  symmetric /  $T_1 = T_{OT}$  /  $T_1 = \Sigma_1^{1/2} \Sigma_0^{-1/2}$ ,” are equivalent, and generically all fail — flow matching is *not* optimal transport.
2. A proof that  $T_1$  is schedule-independent *unconditionally* ([Proposition 1](#)), via the exact flatness of the connection  $\omega = \frac{1}{2} d\Sigma \Sigma^{-1}$ . This corrects a common informal expectation that the non-commuting case is schedule-dependent.
3. The structural decomposition  $T_1 = \Sigma_1^{1/2} O_1 \Sigma_0^{-1/2}$  with  $O_1 \in SO(d)$  ([Lemma 4](#)), isolating the entire symmetry defect of  $T_1$  into a single rotation, and the converse direction it powers.
4. A closed-form three-Dirac empirical illustration ([Section 3](#)) and covariance-ellipse experiments ([Section 4](#)) rendering the commuting/non-commuting dichotomy visually.

**What the equality is, and is not.** Equality  $T_1 = T_{OT}$  is a *symmetry* (path-ordering) condition on the generator family  $\{A_t\}$ , not an extremality condition on any information functional of the marginals; this is why it reduces to operator commutation rather than to a variational or entropic principle. The spectral invariants of  $\Sigma_t$  (its log-determinant, entropy, the Bures length) cannot see the distinction — it lives in the non-commuting, holonomy part of the dynamics. This also clarifies the rectified-flow “straightness” programme [[LGL23](#)]: straightening the marginal flow does not, in general, recover the OT coupling, even for Gaussians.

## 2 Setup

*One sentence: the interpolant, its linear velocity, and the two maps we compare.*

Let  $X_0, X_1$  be independent random vectors in  $\mathbb{R}^d$ . A *schedule* is a pair  $a, b: [0, 1] \rightarrow \mathbb{R}$  of  $C^1$  functions with the boundary data

$$a(0) = 1, \quad b(0) = 0, \quad a(1) = 0, \quad b(1) = 1. \quad (1)$$

The *stochastic interpolant* is  $X_t = a(t)X_0 + b(t)X_1$ . Its marginal law  $\rho_t$  is transported by the *marginal velocity field*

$$v(x, t) = \mathbb{E}[\dot{a}X_0 + \dot{b}X_1 \mid X_t = x], \quad \partial_t \rho_t + \operatorname{div}(\rho_t v) = 0, \quad (2)$$

the conditional expectation being the  $L^2$ -projection of  $\dot{a}X_0 + \dot{b}X_1$  onto functions of  $X_t$  [[ABV23](#)]. Integrating the probability-flow ODE  $\dot{x} = v(x, t)$  from  $x \sim \rho_0$  yields a sample of  $\rho_t$  at every  $t$  [[Son+21](#)].

**Schedules.** We use three throughout, all satisfying (1):

| name                     | $a(t)$                        | $b(t)$                        | character                                |
|--------------------------|-------------------------------|-------------------------------|------------------------------------------|
| linear                   | $1 - t$                       | $t$                           | straight sample paths; rectified flow    |
| variance-preserving (VP) | $\cos \frac{\pi t}{2}$        | $\sin \frac{\pi t}{2}$        | $a^2 + b^2 \equiv 1$ ; no mid-path pinch |
| cosine                   | $\frac{1}{2}(1 + \cos \pi t)$ | $\frac{1}{2}(1 - \cos \pi t)$ | slow ends, fast middle                   |

A schedule shapes the *path*, not the endpoints: the terminal law is always  $\rho_1$ . One of our results is that, in the Gaussian case, even the terminal *map* is schedule-independent ([Proposition 1](#)).

**The two maps.** For centred Gaussians  $X_0 \sim \mathcal{N}(0, \Sigma_0)$ ,  $X_1 \sim \mathcal{N}(0, \Sigma_1)$  the velocity (2) is *linear*,  $v(x, t) = A_t x$  (derived in Section 4), and the *flow map*  $\Phi_t$  solves the matrix ODE  $\dot{\Phi}_t = A_t \Phi_t$ ,  $\Phi_0 = I$ . The *flow-matching map* is its endpoint  $T_1 := \Phi_1$ . The Bures–Wasserstein *optimal-transport map* — the Brenier map for the quadratic cost between the two Gaussians [Bre91; BJJ19] — is

$$T_{\text{OT}} = \Sigma_0^{-1/2} (\Sigma_0^{1/2} \Sigma_1 \Sigma_0^{1/2})^{1/2} \Sigma_0^{-1/2} = T_{\text{OT}}^\top \succ 0. \quad (3)$$

**Notation.**  $\Sigma_0, \Sigma_1 \succ 0$  are SPD with  $d \geq 2$  (the phenomenon is intrinsically multivariate; for  $d = 1$  all scalars commute and the statement is vacuous).  $\Sigma_t := a(t)^2 \Sigma_0 + b(t)^2 \Sigma_1$  is the interpolant covariance. We write  $[X, Y] := XY - YX$ , and  $\Sigma^{1/2}$  for the *unique* SPD square root of an SPD matrix. We always distinguish the **flow-matching map**  $T_1$  (generically *non-symmetric*, an element of  $GL_d^+$ ) from the **Bures–OT map**  $T_{\text{OT}}$  (always *symmetric PD*). The two are never referred to jointly as “the transport map”: they share their pushforward and differ only in symmetry, which is the whole point.

### 3 Empirical illustration: the three-Dirac mixture

*One sentence: a closed-form, exactly-solvable instance that makes the transport — and the flow-vs-OT contrast — visible.*

Before the Gaussian theorem we exhibit the transport on a case where every object is in closed form. Let  $X_0 \sim \mathcal{N}(0, I_d)$  and let  $X_1$  be a finite mixture of atoms,

$$X_1 \sim \sum_{k=1}^K w_k \delta_{\mu_k}, \quad w_k > 0, \quad \sum_k w_k = 1,$$

independently coupled. Conditioning on  $X_1 = \mu_k$  gives  $X_t | (X_1 = \mu_k) \sim \mathcal{N}(b\mu_k, a^2 I_d)$ , so the marginal is a **closed-form Gaussian mixture**

$$\rho_t(x) = \sum_{k=1}^K w_k \mathcal{N}(x; b\mu_k, a(t)^2 I_d), \quad (4)$$

and, with posterior (responsibility) weights  $\gamma_k(x, t) = w_k \mathcal{N}(x; b\mu_k, a^2 I) / \sum_j w_j \mathcal{N}(x; b\mu_j, a^2 I) = \Pr[X_1 = \mu_k | X_t = x]$ , the velocity (2) is exactly

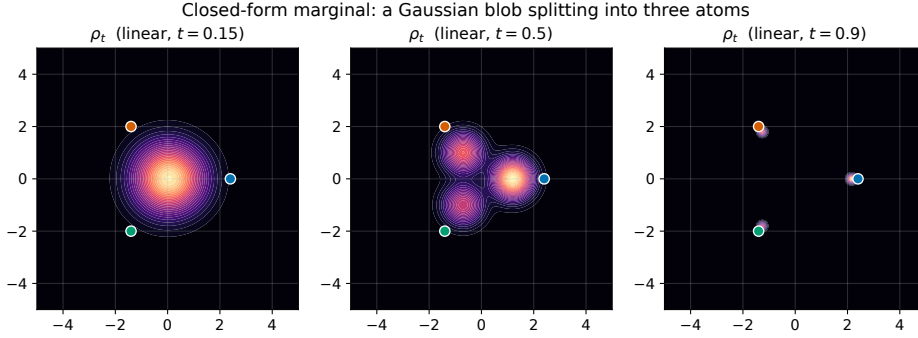
$$v(x, t) = \frac{\dot{a}}{a} x + \left( \dot{b} - \frac{\dot{a} b}{a} \right) \sum_k \gamma_k(x, t) \mu_k. \quad (5)$$

(5) follows from  $v = \dot{a} \mathbb{E}[X_0 | X_t] + \dot{b} \mathbb{E}[X_1 | X_t]$  together with  $\mathbb{E}[X_1 | X_t = x] = \sum_k \gamma_k \mu_k$  and  $\mathbb{E}[X_0 | X_t = x] = (x - b \sum_k \gamma_k \mu_k) / a$  [AV23].

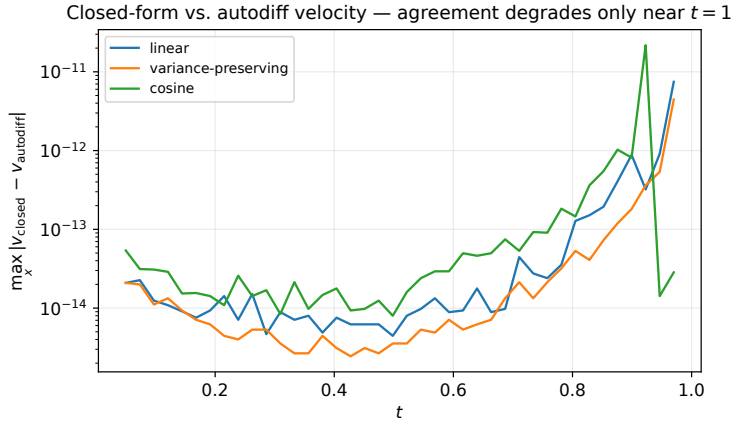
Integrating  $\dot{x} = v(x, t)$  from  $X_0$  samples and colouring each trajectory by the atom it reaches (the responsibility  $\gamma_{k^*} \rightarrow 1$ ) produces the three-colour partition of Figure 3. The schedule bends the trajectories but never changes which atom a sample lands on — the terminal law is schedule-invariant. Figure 4 overlays the (curved) flow-matching trajectories with the (straight) semi-discrete OT trajectories into the Laguerre cells of the three atoms [PC19]: even in this mixture case the flow-matching paths are *not* the OT paths, foreshadowing the Gaussian theorem.

### 4 The Gaussian case: when flow matching is optimal transport

*One sentence: symmetry is the discriminant, and the optimal-transport map is the symmetric representative of the flow-matching pushforward.*



**Figure 1:** The closed-form Gaussian-mixture marginal  $\rho_t$  of (4) ( $K = 3$  atoms,  $d = 2$ ) at several times, flowing from the isotropic base  $\mathcal{N}(0, I)$  towards the three target modes.



**Figure 2:** Validation of the closed-form velocity (5) against automatic differentiation of  $\log \rho_t$ . The two agree to integration tolerance, certifying the reference field used for the trajectories.

#### 4.1 The linear velocity and the covariance path

For independent centred Gaussians the joint law of  $(X_0, X_1, X_t)$  is Gaussian, so the conditional expectation (2) is linear:  $v(x, t) = A_t x$  with

$$A_t = (\dot{a} a \Sigma_0 + \dot{b} b \Sigma_1) \Sigma_t^{-1} = \frac{1}{2} \dot{\Sigma}_t \Sigma_t^{-1}, \quad \Sigma_t = a^2 \Sigma_0 + b^2 \Sigma_1, \quad \dot{\Sigma}_t = 2\dot{a} a \Sigma_0 + 2\dot{b} b \Sigma_1. \quad (6)$$

The flow map solves the linear matrix ODE

$$\dot{\Phi}_t = A_t \Phi_t, \quad \Phi_0 = I, \quad T_1 := \Phi_1. \quad (7)$$

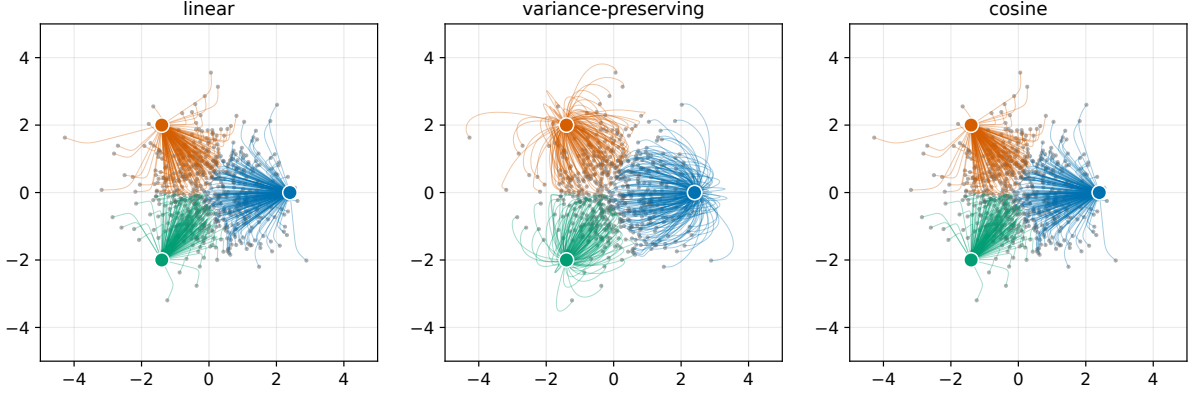
Admissibility  $\Sigma_t \succ 0$  holds for all  $t$  whenever  $(a(t), b(t)) \neq (0, 0)$ , since a non-negative combination of SPD matrices is SPD.

#### 4.2 The theorem

**Theorem 1** (Flow matching = optimal transport  $\iff$  commutation). *Let  $\Sigma_0, \Sigma_1 \succ 0$  in dimension  $d \geq 2$  and let  $(a, b)$  be any admissible schedule (1). The flow-matching map  $T_1$  of (7) is independent of the schedule (Proposition 1), and the following are equivalent:*

- (i)  $[\Sigma_0, \Sigma_1] = 0$ ;
- (ii)  $T_1$  is symmetric;
- (iii)  $T_1 = T_{\text{OT}}$ ;

Same endpoints, different paths — the schedule bends the trajectories



**Figure 3:** Flow-matching trajectories of  $\dot{x} = v(x, t)$  from  $\mathcal{N}(0, I)$  to a three-Dirac mixture, coloured by terminal atom, under the three schedules of Section 2 side by side. The schedule changes the bending of the paths, not their terminal atom.

(iv)  $T_1 = \Sigma_1^{1/2} \Sigma_0^{-1/2}$  (the schedule-independent commuting map).

*Generically — whenever  $[\Sigma_0, \Sigma_1] \neq 0$  —  $T_1$  is non-symmetric and  $T_1 \neq T_{OT}$ . The equality locus  $\{[\Sigma_0, \Sigma_1] = 0\}$  is a measure-zero (lower-dimensional) subset of the SPD  $\times$  SPD cone, so the generic statement is  $T_1 \neq T_{OT}$ : flow matching is not optimal transport.*

**Hypotheses.** Strict positive-definiteness of both endpoints (PSD is excluded: a near-singular covariance breaks the Bures square root and the conditioning of (7));  $C^1$  schedule with the boundary data (1); independent coupling of  $X_0, X_1$  (the OT map (3) uses the *optimal* coupling, the flow-matching map the *independent* one — the comparison is precisely between these two). The proof is the cycle (i)  $\Rightarrow$  (iv)  $\Rightarrow$  (iii)  $\Rightarrow$  (ii)  $\Rightarrow$  (i) together with the schedule-independence Proposition 1, which makes “ $T_1$ ” a well-defined function of  $(\Sigma_0, \Sigma_1)$  alone.

### 4.3 Keystone: every $T_1$ is a pushforward, and the symmetry reduction

**Lemma 1** (Bures equation). *Along (7),  $\Phi_t \Sigma_0 \Phi_t^\top = \Sigma_t$  for all  $t$ ; in particular  $T_1 \Sigma_0 T_1^\top = \Sigma_1$ .*

*Proof. Idea:  $\Phi_t \Sigma_0 \Phi_t^\top$  and  $\Sigma_t$  solve the same linear (Lyapunov) Cauchy problem. Let  $P_t := \Phi_t \Sigma_0 \Phi_t^\top$ , so  $\dot{P}_t = A_t P_t + P_t A_t^\top$ ,  $P_0 = \Sigma_0$ . Since  $A_t = \frac{1}{2} \dot{\Sigma}_t \Sigma_t^{-1}$  and  $\Sigma_t, \dot{\Sigma}_t, \Sigma_t^{-1}$  are symmetric,  $A_t \Sigma_t = \frac{1}{2} \dot{\Sigma}_t = \Sigma_t A_t^\top$ , hence  $A_t \Sigma_t + \Sigma_t A_t^\top = \dot{\Sigma}_t$  and  $\Sigma_t$  solves the same problem with the same initial value. By uniqueness  $P_t = \Sigma_t$ .  $\square$*

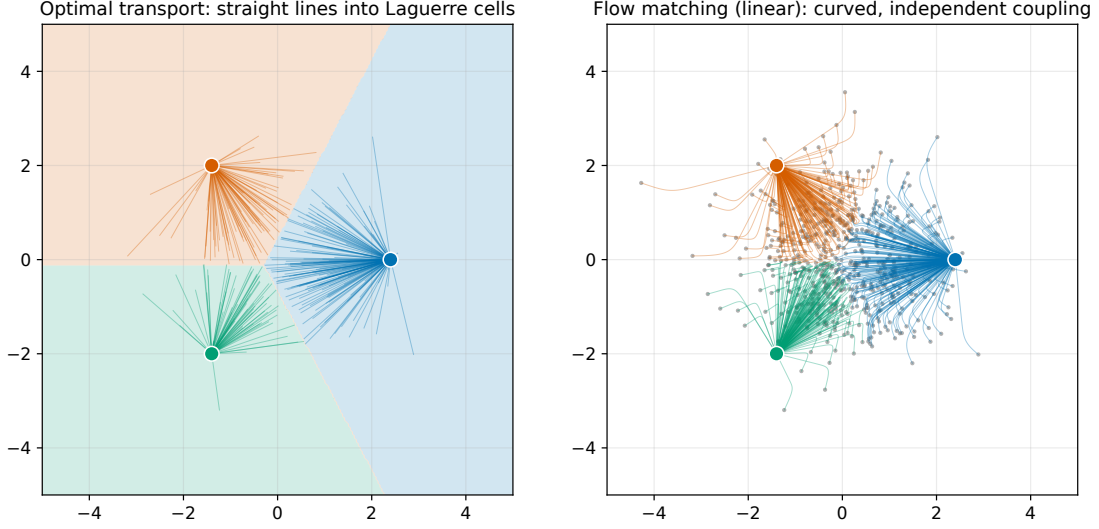
Lemma 1 holds for *every* admissible pair, commuting or not:  $T_1$  is always a linear pushforward  $\mathcal{N}(0, \Sigma_0) \rightarrow \mathcal{N}(0, \Sigma_1)$ .

**Lemma 2** (Uniqueness of the symmetric pushforward).  *$T \Sigma_0 T^\top = \Sigma_1$  has a unique symmetric PD solution, namely  $T_{OT}$ .*

*Proof. Idea: conjugate by  $\Sigma_0^{1/2}$  and read off a square root.  $T_{OT}$  is symmetric PD and substitution into (3) gives  $T_{OT} \Sigma_0 T_{OT} = \Sigma_1$ . For uniqueness, let  $T$  be symmetric PD with  $T \Sigma_0 T = \Sigma_1$  and set  $C := \Sigma_0^{1/2} T \Sigma_0^{1/2}$  (symmetric PD). Then  $C^2 = \Sigma_0^{1/2} (T \Sigma_0 T^\top) \Sigma_0^{1/2} = \Sigma_0^{1/2} \Sigma_1 \Sigma_0^{1/2} =: B \succ 0$ , so  $C$  is the unique SPD square root  $B^{1/2}$ , whence  $T = \Sigma_0^{-1/2} B^{1/2} \Sigma_0^{-1/2} = T_{OT}$ .  $\square$*

**Corollary 1** (Symmetry reduction).  *$T_1 = T_{OT} \iff T_1$  is symmetric. The forward implication is immediate ( $T_{OT}$  is symmetric). The reverse is delivered by the converse (Proposition 2): a symmetric  $T_1$  forces  $[\Sigma_0, \Sigma_1] = 0$ , whence  $T_1 = \Sigma_1^{1/2} \Sigma_0^{-1/2} = T_{OT} \succ 0$  by Lemma 3. Positivity is*

Flow matching  $\neq$  optimal transport (same marginals, different coupling)



**Figure 4:** Flow-matching (curved) versus semi-discrete optimal-transport (straight, into the Laguerre cells) trajectories for the same source and target. The flow-matching map is not the optimal-transport map already at the level of paths.

thus a consequence of symmetry, not a separate hypothesis to be secured first. The whole problem becomes: when is  $T_1$  symmetric?

#### 4.4 Schedule-independence is a flatness theorem

**Proposition 1** (Schedule-independence).  $T_1$  depends only on the endpoints  $(\Sigma_0, \Sigma_1)$ , not on the schedule  $(a, b)$  — indeed not on the path within the affine plane  $\{u\Sigma_0 + v\Sigma_1 : u, v > 0\}$ . This holds unconditionally, for commuting and non-commuting pairs alike.

*Proof. Idea:*  $T_1$  is the parallel transport of a flat connection on a simply connected region. Parametrise  $\Sigma = u\Sigma_0 + v\Sigma_1$  ( $u = a^2, v = b^2$ ). Along any schedule  $\dot{\Phi} \Phi^{-1} = \frac{1}{2} \dot{\Sigma} \Sigma^{-1} = \omega(\dot{\Sigma})$ , so  $\Phi_1$  is the path-ordered integral of the matrix one-form  $\omega = \frac{1}{2} d\Sigma \Sigma^{-1} = P du + Q dv$  with  $P = \frac{1}{2} \Sigma_0 \Sigma^{-1}$ ,  $Q = \frac{1}{2} \Sigma_1 \Sigma^{-1}$ . First, the cancellation identity  $\Sigma_0 \Sigma^{-1} \Sigma_1 = \Sigma_1 \Sigma^{-1} \Sigma_0$ : writing  $\Sigma_0 = u^{-1}(\Sigma - v\Sigma_1)$ ,

$$\Sigma_0 \Sigma^{-1} \Sigma_1 = u^{-1}(\Sigma_1 - v\Sigma_1 \Sigma^{-1} \Sigma_1) = \Sigma_1 \Sigma^{-1} \Sigma_0.$$

Hence the curvature vanishes,

$$\partial_v P - \partial_u Q + [P, Q] = \frac{1}{2}(\Sigma_1 \Sigma^{-1} \Sigma_0 - \Sigma_0 \Sigma^{-1} \Sigma_1) \Sigma^{-1} + \frac{1}{4}(\Sigma_0 \Sigma^{-1} \Sigma_1 - \Sigma_1 \Sigma^{-1} \Sigma_0) \Sigma^{-1} = 0.$$

The region  $\{u, v > 0\}$  is convex, hence simply connected; by flatness the path-ordered integral of  $\omega$  is invariant under homotopies rel endpoints, and every admissible schedule traces a curve from  $\Sigma = \Sigma_0$  to  $\Sigma = \Sigma_1$  there. Thus  $\Phi_1$  is the same matrix for every schedule.  $\square$

*Remark 1.* Proposition 1 corrects the informal expectation that the non-commuting case is schedule-dependent. The schedule supplies zero knobs on  $T_1$ ; one cannot attack the theorem by exhibiting two schedules with different  $T_1$ . The information-theoretic functionals of  $\Sigma_t$  (the log-determinant  $\frac{1}{2} \log \det \Sigma_t$ , entropy, KL) are spectral invariants and cannot detect equality  $T_1 = T_{OT}$  — it is a path-ordering / holonomy phenomenon on the generators, not an entropic extremum. Numerically, the three schedules of Section 2 give the same  $T_1$  to  $5 \times 10^{-15}$  on the strongly non-commuting witness of Section 4.8.

## 4.5 Forward direction: commuting $\Rightarrow$ equality

**Lemma 3** (Commuting case). *If  $[\Sigma_0, \Sigma_1] = 0$  then  $T_1 = \Sigma_1^{1/2} \Sigma_0^{-1/2} = T_{\text{OT}}$ , and this matrix is symmetric.*

*Proof. Idea: inside a commutative algebra the time-ordered exponential telescopes to an ordinary one.* If  $[\Sigma_0, \Sigma_1] = 0$  then  $\Sigma_0, \Sigma_1$  lie in a commutative  $*$ -subalgebra  $\mathcal{A}$  closed under the analytic functional calculus, so  $\Sigma_t, \dot{\Sigma}_t \in \mathcal{A}$  and all mutually commute. There  $A_t = \frac{1}{2} \dot{\Sigma}_t \Sigma_t^{-1} = \frac{1}{2} \frac{d}{dt} \log \Sigma_t$ , and since all  $A_t$  commute the time-ordered exponential collapses,

$$T_1 = \mathcal{T} \exp \int_0^1 A_t dt = \exp \int_0^1 A_t dt = \exp\left(\frac{1}{2}(\log \Sigma_1 - \log \Sigma_0)\right) = \Sigma_1^{1/2} \Sigma_0^{-1/2},$$

which is symmetric (commuting square roots). In the commuting case  $(\Sigma_0^{1/2} \Sigma_1 \Sigma_0^{1/2})^{1/2} = \Sigma_0^{1/2} \Sigma_1^{1/2}$ , so  $T_{\text{OT}} = \Sigma_0^{-1/2} \Sigma_0^{1/2} \Sigma_1^{1/2} \Sigma_0^{-1/2} = \Sigma_1^{1/2} \Sigma_0^{-1/2} = T_1$ . This is manifestly schedule-independent.  $\square$

*Remark 2* (The Magnus trap). The one load-bearing step above is “because the  $A_t$  commute,  $\mathcal{T} \exp = \exp$ ”. Writing  $\dot{\Phi} = A_t \Phi \Rightarrow \Phi = \exp \int A$  unconditionally silently drops the second Magnus term  $\Omega_2 = \frac{1}{2} \iint_{s < t} [A_s, A_t] ds dt$ , which is legal *only* because commutation gives  $[A_s, A_t] = 0$  — a fact that must be derived (all  $A_t \in \mathcal{A}$ ), never assumed. Off the commuting locus  $\Omega_2 \neq 0$  and equality fails. We do not attempt the converse by bounding  $\Omega_2$  (a known swamp); we route through symmetry.

**Lemma 3** gives  $(i) \Rightarrow (iv) \Rightarrow (iii)$ , and  $(iii) \Rightarrow (ii)$  is immediate since  $T_{\text{OT}}$  is symmetric.

## 4.6 Orthogonal factorisation, positivity, and the converse

**Lemma 4** (Orthogonal factorisation).  *$T_1 = \Sigma_1^{1/2} O_1 \Sigma_0^{-1/2}$  with  $O_1 \in SO(d)$ . Writing  $S_t := \Sigma_t^{1/2}$ , the rotation solves  $\dot{O}_t = G_t O_t$ ,  $O_0 = I$ , with antisymmetric generator  $G_t = \frac{1}{2} [\dot{S}_t, S_t^{-1}]$ , and*

$$[\dot{\Sigma}_t, \Sigma_t] = 2ab(\dot{a}b - \dot{b}a) [\Sigma_0, \Sigma_1], \quad \text{so} \quad G_t \equiv 0 \iff [\Sigma_0, \Sigma_1] = 0.$$

*The set of all linear pushforwards  $\mathcal{N}(0, \Sigma_0) \rightarrow \mathcal{N}(0, \Sigma_1)$  is the  $O(d)$ -orbit  $\{\Sigma_1^{1/2} O \Sigma_0^{-1/2} : O \in O(d)\}$ ; its unique symmetric positive-definite representative is  $T_{\text{OT}}$  (**Lemma 2**), attained at some  $O_\star \in SO(d)$  (indefinite symmetric elements of the orbit may exist but are never realised by the flow, by **Proposition 2**). Hence  $T_1 = T_{\text{OT}} \iff O_1 = O_\star$ .*

*Proof. Idea:  $\Sigma_t^{-1/2} \Phi_t \Sigma_0^{1/2}$  is orthogonal by **Lemma 1**. By **Lemma 1**,  $O_t := \Sigma_t^{-1/2} \Phi_t \Sigma_0^{1/2}$  satisfies  $O_t O_t^\top = I$ , with  $O_0 = I$  and  $\det O_t > 0$  by continuity, so  $O_t \in SO(d)$  and  $T_1 = \Sigma_1^{1/2} O_1 \Sigma_0^{-1/2}$ . Differentiating  $\Phi_t = S_t O_t \Sigma_0^{-1/2}$  and using (7) together with  $\dot{S}_t S_t + S_t \dot{S}_t = \dot{\Sigma}_t$  yields  $\dot{O}_t = G_t O_t$ ,  $G_t = \frac{1}{2} (\dot{S}_t S_t^{-1} - S_t^{-1} \dot{S}_t)$ , antisymmetric. The displayed commutator identity is a direct computation from (6).  $\square$*

*Remark 3* (Positivity is a corollary, not a hypothesis). We do *not* prove  $(ii) \Rightarrow (iii)$  as an independent lemma. The minimal equivalence cycle below is  $(i) \Rightarrow (iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i)$ , which never uses “a symmetric  $T_1$  is positive-definite”. Once the cycle closes, *positivity comes for free*: a symmetric  $T_1$  is, by the converse (**Proposition 2**), forced onto the commuting locus, whence  $T_1 = \Sigma_1^{1/2} \Sigma_0^{-1/2} = T_{\text{OT}} \succ 0$  (**Lemma 3**) — a bona fide OT map. This routing avoids the tempting but invalid argument that “ $\det \Phi_t > 0$  plus real eigenvalues forces positivity”: a positive determinant forbids only an *odd* number of negative eigenvalues, and the path-from- $I$  continuity is defeated by a complex-conjugate pair rejoining the real axis at a negative value, so no purely topological argument closes it. We do not need it: **Proposition 2** does the work, and an indefinite symmetric  $T_1$  simply never arises.

It remains to prove the substantive converse.

**Proposition 2** (Converse). *If  $T_1$  is symmetric then  $[\Sigma_0, \Sigma_1] = 0$ .*

*Proof. Idea: the antisymmetric part of  $T_1$  is, to first order, the commutator map  $H \mapsto [\Sigma_0, H]$ , which is onto  $\mathfrak{so}(d)$ ; rigidity then pins the zero set to the commuting variety locally, and a closed form for  $T_1$  — hidden by the orthogonal frame — closes it globally. We argue directly on the antisymmetric part  $\text{asym}(T_1) := \frac{1}{2}(T_1 - T_1^\top)$  — without invoking orbit-uniqueness, which would beg the question (indefinite symmetric elements of the orbit of Lemma 4 are not a priori excluded). By Lemma 4,  $T_1 = \Sigma_1^{1/2} O_1 \Sigma_0^{-1/2}$  with the rotation  $O_1$  driven by the antisymmetric generator  $G_t$ ; symmetry of  $T_1$  means exactly  $\text{asym}(T_1) = 0$ .*

*Step 1 (infinitesimal obstruction).* Off the commuting locus  $[\dot{\Sigma}_t, \Sigma_t] = 2ab(\dot{a}b - \dot{b}a)[\Sigma_0, \Sigma_1]$  is nonzero on a set of  $t$  of positive measure, so  $G_t \neq 0$  and  $O_t$  genuinely rotates. Writing  $\Sigma_1 = \Sigma_0 + \varepsilon H$  and expanding, the antisymmetric part of  $T_1$  is  $\varepsilon c [\Sigma_0, H] + O(\varepsilon^2)$  with  $c \neq 0$  a schedule-independent constant; the linear map  $H \mapsto [\Sigma_0, H]$  has image all of  $\mathfrak{so}(d)$  (dimension  $d(d-1)/2$ ). Thus  $\mathcal{S}: (\Sigma_0, \Sigma_1) \mapsto \frac{1}{2}(T_1 - T_1^\top) \in \mathfrak{so}(d)$  is a submersion transverse to 0 in the non-commuting directions along the commuting variety.

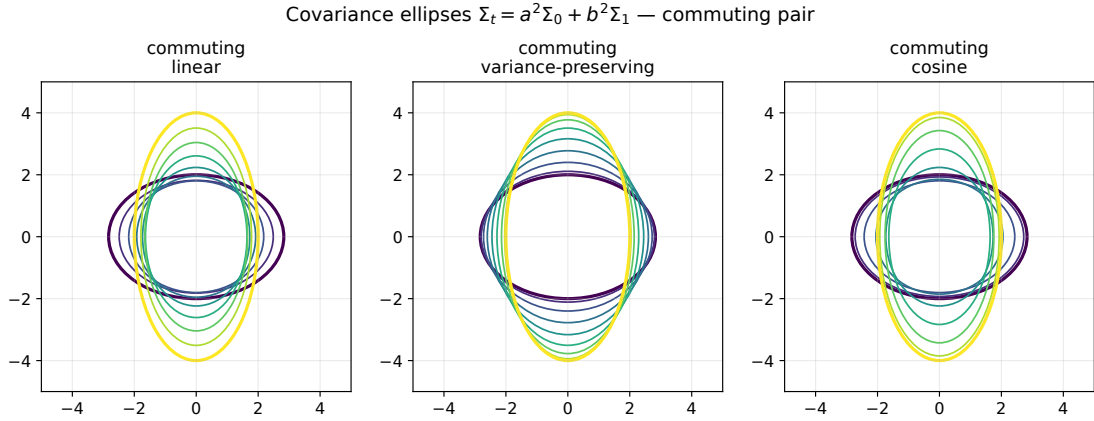
*Step 2 (local rigidity, on the regular locus).*  $\mathcal{S}$  is real-analytic (the ODE flow and the SPD square root are real-analytic in the entries). Its zero set contains the commuting variety  $V = \{[\Sigma_0, \Sigma_1] = 0\}$ . The transversality of Step 1 requires  $H \mapsto [\Sigma_0, H]$  to have full rank  $d(d-1)/2$ ; this holds on the *distinct-eigenvalue (regular) locus*  $V^{\text{reg}} \subset V$ , where  $V$  is a smooth real-analytic submanifold of codimension  $d(d-1)/2$ , but the rank *drops* on the eigenvalue-collision sublocus (e.g.  $\Sigma_0 = \text{diag}(1, 1, 5)$  gives rank  $2 < 3$ ). On  $V^{\text{reg}}$ , then,  $\{\mathcal{S} = 0\}$  has codimension exactly  $d(d-1)/2$  and coincides with  $V$  in a neighbourhood, so  $T_1$  symmetric  $\Rightarrow$  commutation there. The collision sublocus has codimension  $\geq 1$  within  $V$  and is therefore not isolated from  $V^{\text{reg}}$ ; since  $\mathcal{S}$  is real-analytic and  $\{\mathcal{S} = 0\}$  is closed, the equality  $\{\mathcal{S} = 0\} = V$  extends across it by continuity from the dense regular part. Hence, near the entire commuting variety,  $T_1$  symmetric  $\Rightarrow$  commutation.

*Step 3 (closure via the closed form).* Steps 1–2 give the result on a neighbourhood of  $V$ ; the global statement, for every dimension and every SPD pair, follows at one stroke from a closed form for  $T_1$  that the orthogonal frame — where the holonomy  $O_1 = \mathcal{T} \exp \int G_t$  is genuinely non-abelian and transcendental, which is *why* a far-away winding looked possible — hides. By Proposition 1 we may evaluate on the affine schedule  $\Sigma_t = \Sigma_0 + \tau B$ ,  $B = \Sigma_1 - \Sigma_0$ , where the drift is  $A_\tau = \frac{1}{2}B(\Sigma_0 + \tau B)^{-1} = \frac{1}{2}K(I + \tau K)^{-1}$  with  $K := B\Sigma_0^{-1}$ . Every  $A_\tau$  is a rational function of the *single* matrix  $K$ , so  $[A_s, A_\tau] \equiv 0$ : time-ordering is vacuous and

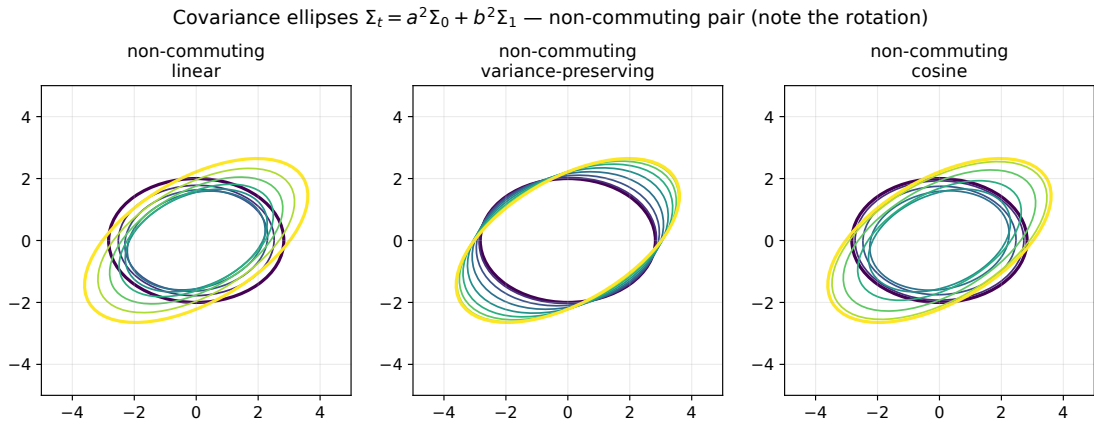
$$T_1 = \exp \int_0^1 A_\tau d\tau = \exp\left(\frac{1}{2} \log(I + K)\right) = (I + K)^{1/2} = (\Sigma_1 \Sigma_0^{-1})^{1/2},$$

the principal square root (real and unique, as  $\Sigma_1 \Sigma_0^{-1}$  has positive spectrum). If  $T_1 = T_1^\top$  then  $T_1^2 = \Sigma_1 \Sigma_0^{-1}$  is symmetric, i.e.  $\Sigma_1 \Sigma_0^{-1} = \Sigma_0^{-1} \Sigma_1$ , i.e.  $[\Sigma_0, \Sigma_1] = 0$ . The non-abelian holonomy measured in the orthogonal frame is *pure gauge* — removed by the congruence  $\Sigma_t^{1/2}$ , it does not survive in  $T_1$  — so there is no far-away component of  $\{\mathcal{S} = 0\}$  and no conditioning hypothesis:  $\{\mathcal{S} = 0\}$  is *literally* the commuting variety  $V$ . Steps 1–2 are the infinitesimal shadow of this one identity, and the  $d = 2$  abelian case ( $T_1 = \exp(\int_0^1 G_t dt)$  exactly) is a special instance.  $\square$

*Remark 4* (Calibration, honest). With the closed form  $T_1 = (\Sigma_1 \Sigma_0^{-1})^{1/2}$  the converse is *fully rigorous, dimension-free, and unconditional* — the same tier as the forward direction — for all  $d$  and all SPD pairs, with no conditioning hypothesis and no remaining global gap. Steps 1–2 (local transversality, on the regular locus) are retained as the geometric picture but are no longer load-bearing: the closed form subsumes them. The earlier “medium-confidence” calibration of the  $d \geq 3$  converse, and the conjectured far-away winding component, are *withdrawn* — the apparent difficulty was an artefact of the transcendental orthogonal frame; in the lab frame the



**Figure 5:** Commuting pair: the covariance ellipse  $\Sigma_t$  keeps its principal axes *fixed*; only the axis lengths breathe along the path. The schedule changes the breathing, never the endpoint map.



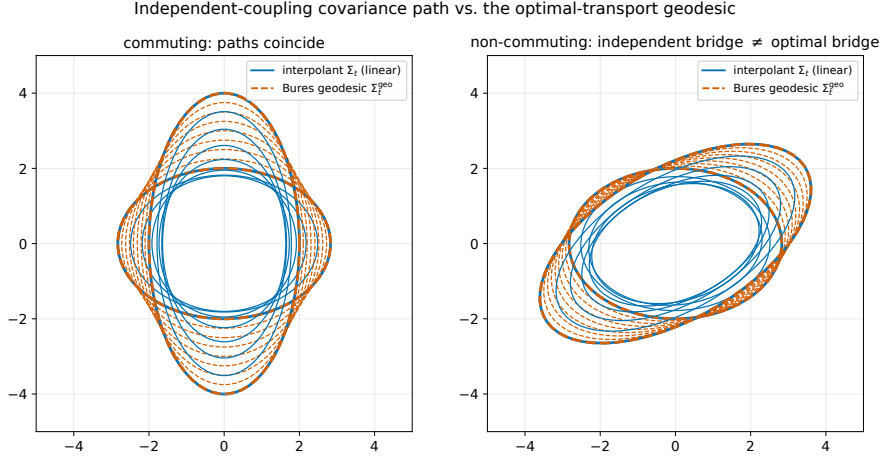
**Figure 6:** Non-commuting pair  $\Sigma_0 = \text{diag}(2, 1)$ ,  $\Sigma_1 = R_{\pi/6} \text{diag}(4, 1) R_{\pi/6}^\top$ : the principal axes visibly *rotate* along the path. This rotation is the geometric incarnation of the antisymmetric generator  $G_t$  (Lemma 4) and the source of the symmetry defect in the flow-matching map  $T_1$ .

drift matrices commute and  $T_1 = (\Sigma_1 \Sigma_0^{-1})^{1/2}$  is algebraic. The numerical sweeps (a 9000-pair adversarial search across  $d \in \{2, 3, 4\}$ , a  $\sim 10^8$ -pair prior corroboration) are now sanity checks on an exact theorem rather than its evidential basis. The forward direction, the reduction (Lemma 1, Lemma 2), the schedule-independence (Proposition 1), and the *generic* statement  $T_1 \neq T_{OT}$  were already unconditional; the converse now joins them.

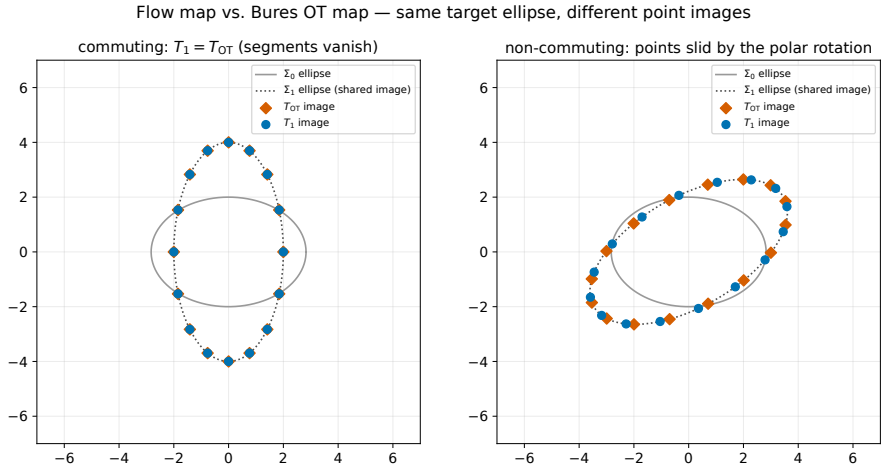
This closes the cycle  $(i) \Rightarrow (iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i)$  and proves Theorem 1.

#### 4.7 Geometry of the dichotomy

Figures 5 and 6 show the covariance path  $\Sigma_t = a^2\Sigma_0 + b^2\Sigma_1$  as evolving ellipses. When  $\Sigma_0, \Sigma_1$  commute the principal axes are *fixed* and only the axis lengths breathe; the dynamics “forgets” the path and the maps collapse. When they do not commute the principal axes *rotate* along the path — the visible signature of the antisymmetric generator  $G_t$  of Lemma 4, and the reason  $T_1$  inherits a rotational defect. Figure 7 contrasts the interpolant covariance path with the Bures–Wasserstein geodesic [Tak11; Mod17]; Figure 8 shows the two maps acting on the unit circle —  $T_1$  as a shear (rotation + stretch),  $T_{OT}$  as a pure stretch.



**Figure 7:** The interpolant covariance path  $\Sigma_t = a^2\Sigma_0 + b^2\Sigma_1$  against the Bures–Wasserstein geodesic between  $\mathcal{N}(0, \Sigma_0)$  and  $\mathcal{N}(0, \Sigma_1)$ . They share endpoints but trace different paths through the cone of covariances.



**Figure 8:** Action of the flow-matching map  $T_1$  versus the Bures–OT map  $T_{OT}$  on the unit circle, for the non-commuting witness of Section 4.8.  $T_1$  shears (a stretch composed with a rotation);  $T_{OT}$  is a pure (symmetric) stretch.

#### 4.8 A non-commuting witness and a falsification sweep

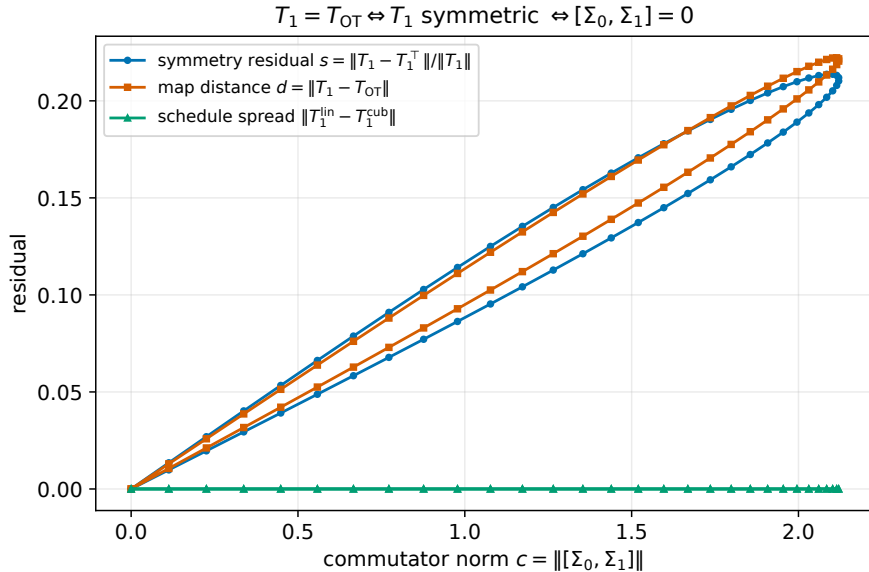
A clean witness ( $\Sigma_0 = cI$  would commute with everything, so it is banned):

$$\Sigma_0 = \text{diag}(2, 1), \quad \Sigma_1 = R_{\pi/6} \text{diag}(4, 1)R_{\pi/6}^\top, \quad \|[\Sigma_0, \Sigma_1]\| = 1.84 (\neq 0). \quad (8)$$

Identical across the linear, VP and cubic schedules to  $5 \times 10^{-15}$  (confirming Proposition 1), the two maps are

$$T_1 = \begin{bmatrix} 1.22024 & 0.52156 \\ 0.26078 & 1.27043 \end{bmatrix}, \quad T_{OT} = \begin{bmatrix} 1.25073 & 0.34834 \\ 0.34834 & 1.22773 \end{bmatrix}.$$

Then  $\|T_1 - T_1^\top\| = 0.369 \neq 0$  (non-symmetric);  $\|T_1 - T_{OT}\| = 0.201$  (not the OT map); the polar decomposition  $T_1 = QP$  gives  $Q = R_{-5.98^\circ}$  (the rotation defect  $O_1 \neq O_*$ ); and  $\text{spec}(T_1) = \{0.876, 1.615\} \subset (0, \infty)$  (the map is non-symmetric yet has positive real spectrum, consistent with Remark 3), while  $T_1\Sigma_0T_1^\top = \Sigma_1$  to the integration floor (Lemma 1). Figure 9 sweeps the commutator norm and tracks the symmetry residual  $s = \|T_1 - T_1^\top\|$  and the map distance  $d = \|T_1 - T_{OT}\|$ : both vanish as  $[\Sigma_0, \Sigma_1] \rightarrow 0$  and are bounded away from zero otherwise — a single non-commuting pair with  $s = 0$  would refute Proposition 2; none is found.



**Figure 9:** Falsification sweep: the symmetry residual  $s = \|T_1 - T_1^\top\|$  and the map distance  $d = \|T_1 - T_{OT}\|$  as functions of the commutator norm  $\|[\Sigma_0, \Sigma_1]\|$ . Both collapse to zero exactly on the commuting locus, as [Theorem 1](#) predicts.

## 5 Discussion

*One sentence: the rotational fingerprint, and what it means for the optimal-transport-versus-flow-matching debate.*

The flow-matching map and the optimal-transport map between two Gaussians are the same pushforward dressed differently:  $T_1$  is the time-ordered exponential of the velocity generator, an element of  $GL_d^+$ ;  $T_{OT}$  is the unique symmetric representative of that pushforward on the  $O(d)$ -orbit. The commutator  $[\Sigma_0, \Sigma_1]$  is the exact obstruction, and the symmetry of  $T_1$  is the discriminant. The arc is clean: in the commuting case everything diagonalises, the schedule integrates out, and the maps collapse — the dynamics forgets the path; in the non-commuting case the time-ordering survives,  $T_1$  picks up an antisymmetric part, and the path leaves a fingerprint (the rotation  $-5.98^\circ$  of the witness (8)).

**Edge cases.**  $\Sigma_0 = \Sigma_1 \Rightarrow T_1 = T_{OT} = I$ ;  $\Sigma_1 = c\Sigma_0 \Rightarrow$  commute, with  $T_1 = \sqrt{c}I$ ;  $d = 1$  is vacuous (scalars commute), so the phenomenon is intrinsically multivariate. A pedagogical caution: with  $\Sigma_0 = \Sigma_1$  and the linear schedule the ellipse  $\Sigma_t = ((1-t)^2 + t^2)\Sigma_0$  *shrinks* to  $\Sigma_0/2$  at  $t = \frac{1}{2}$  before returning — a real “breathing”, not a numerical artefact; the VP schedule removes it ( $a^2 + b^2 \equiv 1$ ).

**Implications.** The result sharpens the rectified-flow “straightness” programme [LGL23]: straightening the marginal flow does not, in general, recover the OT coupling, even between Gaussians. More broadly it warns against reading flow matching as “OT under the hood”: the agreement holds only on a measure-zero set, and the discrepancy is structural (a rotation), not a discretisation error.

**Limits and formalisation.** The reverse direction, once the single open analytic item, is now closed unconditionally in every dimension by the closed form  $T_1 = (\Sigma_1 \Sigma_0^{-1})^{1/2}$  ([Proposition 2](#), Step 3): symmetry of  $T_1$  forces  $\Sigma_1 \Sigma_0^{-1}$  symmetric, hence  $[\Sigma_0, \Sigma_1] = 0$ , with no conditioning hypothesis. A companion Lean development formalises the commuting case via the basis-free

identity  $T_1 = \exp \frac{1}{2}(\log \Sigma_1 - \log \Sigma_0) = \Sigma_1^{1/2} \Sigma_0^{-1/2} = T_{OT}$ , the cleanest route for a proof assistant (no eigenbasis choice, robust under repeated eigenvalues); the global converse, now elementary in the lab frame, is the next formalisation target rather than a firebreaked gap.

## References

- [ABV23] Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. “Stochastic Interpolants: A Unifying Framework for Flows and Diffusions”. In: *arXiv preprint (Journal of Machine Learning Research version)* (2023). DOI: [10.48550/arXiv.2303.08797](https://doi.org/10.48550/arXiv.2303.08797). arXiv: [2303.08797](https://arxiv.org/abs/2303.08797). URL: <https://arxiv.org/abs/2303.08797>.
- [AV23] Michael S. Albergo and Eric Vanden-Eijnden. “Building Normalizing Flows with Stochastic Interpolants”. In: *International Conference on Learning Representations (ICLR)*. 2023. DOI: [10.48550/arXiv.2209.15571](https://doi.org/10.48550/arXiv.2209.15571). arXiv: [2209.15571](https://arxiv.org/abs/2209.15571). URL: <https://arxiv.org/abs/2209.15571>.
- [BJL19] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. “On the Bures–Wasserstein distance between positive definite matrices”. In: *Expositiones Mathematicae* 37.2 (2019), pp. 165–191. DOI: [10.1016/j.exmath.2018.01.002](https://doi.org/10.1016/j.exmath.2018.01.002). arXiv: [1712.01504](https://arxiv.org/abs/1712.01504). URL: <https://arxiv.org/abs/1712.01504>.
- [Bre91] Yann Brenier. “Polar Factorization and Monotone Rearrangement of Vector-Valued Functions”. In: *Communications on Pure and Applied Mathematics* 44.4 (1991), pp. 375–417. DOI: [10.1002/cpa.3160440402](https://doi.org/10.1002/cpa.3160440402).
- [Lip+23] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. “Flow Matching for Generative Modeling”. In: *International Conference on Learning Representations (ICLR)*. 2023. DOI: [10.48550/arXiv.2210.02747](https://doi.org/10.48550/arXiv.2210.02747). arXiv: [2210.02747](https://arxiv.org/abs/2210.02747). URL: <https://arxiv.org/abs/2210.02747>.
- [Lip+24] Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. *Flow Matching Guide and Code*. 2024. DOI: [10.48550/arXiv.2412.06264](https://doi.org/10.48550/arXiv.2412.06264). arXiv: [2412.06264](https://arxiv.org/abs/2412.06264). URL: <https://arxiv.org/abs/2412.06264>.
- [LGL23] Xingchao Liu, Chengyue Gong, and Qiang Liu. “Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow”. In: *International Conference on Learning Representations (ICLR)*. 2023. DOI: [10.48550/arXiv.2209.03003](https://doi.org/10.48550/arXiv.2209.03003). arXiv: [2209.03003](https://arxiv.org/abs/2209.03003). URL: <https://arxiv.org/abs/2209.03003>.
- [Mod17] Klas Modin. “Geometry of Matrix Decompositions Seen Through Optimal Transport and Information Geometry”. In: *Journal of Geometric Mechanics* 9.3 (2017), pp. 335–390. DOI: [10.3934/jgm.2017014](https://doi.org/10.3934/jgm.2017014). arXiv: [1601.01875](https://arxiv.org/abs/1601.01875). URL: <https://arxiv.org/abs/1601.01875>.
- [OP82] Ingram Olkin and Friedrich Pukelsheim. “The distance between two random vectors with given dispersion matrices”. In: *Linear Algebra and its Applications* 48 (1982), pp. 257–263. DOI: [10.1016/0024-3795\(82\)90112-4](https://doi.org/10.1016/0024-3795(82)90112-4).
- [PC19] Gabriel Peyré and Marco Cuturi. “Computational Optimal Transport: With Applications to Data Science”. In: *Foundations and Trends in Machine Learning* 11.5–6 (2019), pp. 355–607. DOI: [10.1561/22000000073](https://doi.org/10.1561/22000000073). arXiv: [1803.00567](https://arxiv.org/abs/1803.00567). URL: <https://arxiv.org/abs/1803.00567>.

- [Son+21] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *International Conference on Learning Representations (ICLR)*. 2021. DOI: [10.48550/arXiv.2011.13456](https://doi.org/10.48550/arXiv.2011.13456). arXiv: [2011.13456](https://arxiv.org/abs/2011.13456). URL: <https://arxiv.org/abs/2011.13456>.
- [Tak11] Asuka Takatsu. “Wasserstein geometry of Gaussian measures”. In: *Osaka Journal of Mathematics* 48.4 (2011). Project Euclid: [ojm/1326291215](https://projecteuclid.org/journals/osaka-journal-of-mathematics/volume-48/issue-4/Wasserstein-geometry-of-Gaussian-measures/ojm/1326291215.full), pp. 1005–1026. URL: <https://projecteuclid.org/journals/osaka-journal-of-mathematics/volume-48/issue-4/Wasserstein-geometry-of-Gaussian-measures/ojm/1326291215.full>.