CONVERGENCE OF THE VERTICAL GRADIENT FLOW FOR THE GAUSSIAN MONGE PROBLEM

Citation for the original published paper (version of record):

Jansson, E., Modin, K. (2024). CONVERGENCE OF THE VERTICAL GRADIENT FLOW FOR THE GAUSSIAN MONGE PROBLEM. Journal of Computational Dynamics, 11(1): 1-9.
http://dx.doi.org/10.3934/jcd.2023008

Journal of Computational Dynamics
Vol. 11, No. 1, January 2024, pp. 1-9
doi:10.3934/jcd.2023008

CONVERGENCE OF THE VERTICAL GRADIENT FLOW FOR THE GAUSSIAN MONGE PROBLEM

Erik Jansson* and Klas Modin

Department of Mathematical Sciences,
Chalmers University of Technology & University of Gothenburg, Sweden

Abstract. We investigate a matrix dynamical system related to optimal mass transport in the linear category, namely, the problem of finding an optimal invertible matrix by which two covariance matrices are congruent. We first review the differential geometric structure of the problem in terms of a principal fiber bundle. The dynamical system is a gradient flow restricted to the fibers of the bundle. We prove global existence of solutions to the flow, with convergence to the polar decomposition of the matrix given as initial data. The convergence is illustrated in a numerical example.

1. Introduction.

Matrix decompositions are ubiquitous in mathematics. An interesting view-point is to think of matrix decompositions as originating from limits of matrix flows. Famous examples include the Toda flow, as described by Flaschka [6], and Brockett's diagonalizing flow [5]. A review of these and many other examples is given by Modin [11], where emphasis is put on the connection to optimal mass transport and information geometry. The approach is based on Riemannian gradient flows, which reveals connections to optimal transport and information geometry. Further, it highlights hidden dynamical features of matrix factorizations, notably that the polar, QR, and singular value decompositions can be obtained via gradient flows.

In this paper, we study the polar decomposition of matrices from a geometric perspective. We show, in particular, that it is given as the limit of the *vertical gradient flow* introduced by Modin [11]. Thus, the contribution of the paper is a careful analysis of this finite-dimensional vertical gradient flow. Our main result is a proof of convergence to a minimizer. The convergence of this flow is interesting in itself, as it completes the picture in Modin [11], where its "cousin" flow – the *horizontal gradient flow* – is shown to converge. But perhaps more interestingly, the result suggests a similar convergence study in the infinite-dimensional case, for the gradient flow corresponding to Brenier's polar decomposition of maps [4], by using the corresponding infinite-dimensional geometric framework of gradient flows developed by Balehowsky, Karlsson, and Modin [2].

2020 *Mathematics Subject Classification.* Primary: 15A23, 49M99; Secondary: 49Q22.

Key words and phrases. Gradient flows, Matrix decompositions, Optimal transport.

This work was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. This work was also supported by the Swedish Research Council, grant number 2022-03453, and the Knut and Alice Wallenberg Foundation, grant number WAF2019.0201.

*Corresponding author: Erik Jansson.

2. **The Gaussian Monge problem.** In this section, we briefly describe the theoretical background, state the optimization problem at the center of our study and outline its geometric structure. For details, see Modin [11] and references therein.

Let μ_0 and μ_1 be zero-mean multivariate Gaussian distributions on $\mathbb{R}^n$. The *Gaussian Monge problem* consists of finding an invertible linear map $\phi(x) = Ax$, with $A \in \mathrm{GL}(n)$ an invertible $n \times n$ real matrix, that pushes μ_0 forward to μ_1 while minimizing the functional $J\colon \mathrm{GL}(n) \to \mathbb{R}$,

$$J(A) := \int_{\mathbb{R}^n} \|x - Ax\|^2 \mu_0. \tag{1}$$

Here, $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^n$.

It is possible to cast the Gaussian Monge problem only in terms of matrices. Indeed, from the perspective of information geometry (cf. Amari [1]), the set of all zero-mean multivariate Gaussian distributions is a *statistical manifold* isomorphic to the set of covariance matrices, i.e., the set $\mathrm{P}(n)$ of symmetric and positive-definite $n \times n$ real matrices.

We identify μ_0 and μ_1 with their covariance matrices, Σ_0 and Σ_1, and note that the functional in equation (1) can be rewritten as

$$J(A) = \mathrm{Tr}(\Sigma_0 (I - A)^T (I - A)), \tag{2}$$

where Tr denotes the matrix trace and I is the identity matrix. The matrix A transforms $\mathbb{R}^n$ and thereby it transforms μ_0. Under the identification of Gaussian distributions with covariance matrices, A maps Σ_0 to $A\Sigma_0 A^T$. Therefore, the Monge problem for multivariate Gaussian distributions can be formulated in terms of covariance matrices.

Problem 1 (Gaussian Monge problem)**.**

$$\begin{aligned} &\min_{A \in \mathrm{GL}(n)} J(A) = \mathrm{Tr}(\Sigma_0 (I - A)^T (I - A)) \\ &\textit{subject to: } A \in \mathcal{C}(\Sigma_0, \Sigma_1) = \{A \in \mathrm{GL}(n) \mid A\Sigma_0 A^T = \Sigma_1\} \end{aligned} \tag{3}$$

To describe the Riemannian geometry of the problem, we begin by equipping $\mathrm{GL}(n)$ with a Riemannian metric $\mathcal{G}$ defined by

$$\mathcal{G}_A(\dot{A}, \dot{A}) = \mathrm{Tr}(\Sigma_0 \dot{A}^T \dot{A}) = \mathrm{Tr}(A\Sigma_0 A^T (\dot{A}A^{-1})^T (\dot{A}A^{-1})) \tag{4}$$

for $\dot{A} \in T_A \mathrm{GL}(n)$. The corresponding Riemannian distance function $d\colon \mathrm{GL}(n) \times \mathrm{GL}(n) \mapsto \mathbb{R}$ is

$$d(A_0, A_1) = \mathrm{Tr}\left(\Sigma_0 (A_0 - A_1)^T (A_0 - A_1)\right)^{1/2},$$

so we see that $J(A) = d^2(I, A)$.

The Gaussian Monge problem is solved by the polar decomposition of matrices, as pointed out by Brenier [4].

Theorem 2.1. *Let $A \in \mathrm{GL}(n)$ and $\Sigma_0 \in \mathrm{P}(n)$. Then there exists unique matrices $P \in \mathrm{P}(n)$ and $Q \in \mathrm{O}(n, \Sigma_0) := \{Q \in \mathrm{GL}(n) \mid Q\Sigma_0 Q^T = \Sigma_0\}$ such that*

$$A = PQ.$$

Moreover, the matrix P is the unique solution to Problem 1 when $\Sigma_1 = \pi(A) = A\Sigma_0 A^T$.

2.1. **The geometric structure of the Gaussian Monge problem.** We now describe the geometric structure of Problem 1. First, $\mathrm{O}(n, \Sigma_0)$ is a Lie subgroup of $\mathrm{GL}(n)$ with Lie algebra

$$\mathfrak{o}(n, \Sigma_0) = \{X \in \mathfrak{gl}(n) : X\Sigma_0 + \Sigma_0 X^T = 0\}.$$

A matrix $A \in \mathrm{GL}(n)$ acts on $\Sigma \in P(n)$ by the matrix congruence action $A.\Sigma := A\Sigma A^T$. This transitive action defines a projection $\pi\colon \mathrm{GL}(n) \to P(n)$ given by

$$\pi(A) = A\Sigma_0 A^T.$$

The isotropy subgroup of Σ_0 with respect to the congruence action is $\mathrm{O}(n, \Sigma_0)$. Further, the Riemannian metric $\mathcal{G}$ on $\mathrm{GL}(n)$ is $\mathrm{O}(n, \Sigma_0)$-invariant. That is, $\mathcal{G}$ is invariant under right multiplication with matrices in $\mathrm{O}(n, \Sigma_0)$. Indeed, with an arbitrary $Q \in \mathrm{O}(n, \Sigma_0)$,

$$\begin{aligned}\mathcal{G}_A(\dot{A}, \dot{A}) &= \mathrm{Tr}(\Sigma_0 \dot{A}^T \dot{A}) = \mathrm{Tr}(Q\Sigma_0 Q^T \dot{A}^T \dot{A}) \\ &= \mathrm{Tr}(\Sigma_0 (\dot{A}Q)^T \dot{A}Q) = \mathcal{G}_{AQ}(\dot{A}Q, \dot{A}Q).\end{aligned}$$

For any $A \in \mathcal{C}(\Sigma_0, \Sigma_1)$, the mapping

$$\mathcal{C}(\Sigma_0, \Sigma_1) \ni B \mapsto A^{-1}B \in \mathrm{O}(n, \Sigma_0) \tag{5}$$

provides an isomorphism between $\mathcal{C}(\Sigma_0, \Sigma_1)$ and $\mathrm{O}(n, \Sigma_0)$. To see this, note that $\pi(AQ) = \Sigma_1$ for all $Q \in \mathrm{O}(n, \Sigma_0)$, and for any $B \in \mathcal{C}(\Sigma_0, \Sigma_1)$, $A^{-1}B \in \mathrm{O}(n, \Sigma_0)$ since $A^{-1}B\Sigma_0 B^T A^{-T} = A^{-1}\Sigma_1 A^{-T} = \Sigma_0$. Furthermore, we have that $\mathcal{C}(\Sigma_0, \Sigma_1) = \pi^{-1}(\Sigma_1)$. The sets $\pi^{-1}(\Sigma)$ are fibers of a principal $\mathrm{O}(n, \Sigma_0)$-bundle over $P(n)$:

$$\begin{array}{ccc} \mathrm{GL}(n) & \hookleftarrow & \mathrm{O}(n, \Sigma_0) \\ \downarrow & & \\ P(n) & & \end{array} \tag{6}$$

The principal bundle structure (6) gives rise to two linear subspaces of each tangent space $T_A\,\mathrm{GL}(n)$. The *vertical distribution* Ver_A is geometrically the tangent space of the fiber going through A, and is determined by computing the kernel of the derivative of $\pi(A)$, denoted $D\pi(A)$. The *horizontal distribution* Hor_A is the distribution transversal to Ver_A with respect to the metric $\mathcal{G}$. Note that the definition of Ver_A relies solely on the bundle structure. Therefore, Ver_A is independent of the choice of metric.

To compute Ver_A we first compute $D\pi(A)$. It is, with $\Sigma = \pi(A)$, given by

$$\dot{\Sigma} = \dot{A}A^{-1}A\Sigma_0 A^T + A\Sigma_0 A^T(\dot{A}A^{-1})^T = \dot{A}A^{-1}\Sigma + \Sigma(\dot{A}A^{-1})^T := D\pi(A)\cdot\dot{A}.$$

Note that $\dot{A}A^{-1} \in \mathfrak{gl}(n)$. A matrix $VA \in T_A\,\mathrm{GL}(n)$ is thus in the kernel of $D\pi(A)$ if V satisfies

$$V\Sigma + \Sigma V^T = 0,$$

which is equivalent to the condition $V \in \mathfrak{o}(n, \Sigma)$. Thus, the tangent space, called the *vertical distribution* at A, is given by

$$\mathrm{Ver}_A = \{\dot{A} = VA \in T_A\,\mathrm{GL}(n) \mid V \in \mathfrak{o}(n, \Sigma)\}.$$

To compute Hor_A, we must compute the orthogonal complement of $\mathfrak{o}(n, \Sigma)$ with respect to $\mathcal{G}$. To this end, let $U \in \mathfrak{gl}(n)$ be arbitrary and note that if $V \in \mathfrak{o}(n, \Sigma)$, then $V\Sigma = X$ is skew-symmetric, so $V = X\Sigma^{-1}$.

Inserting U and V into Equation (4) we obtain

$$\mathcal{G}_A(VA, UA) = \operatorname{Tr}(A\Sigma_0 A^T V^T U) = \operatorname{Tr}(\Sigma V^T U) = \operatorname{Tr}(X^T U).$$

The orthogonal complement of the skew-symmetric matrices under the Frobenius inner product (i.e, $\mathcal{G}$ with $\Sigma_0 = I$), is the set of symmetric matrices, denoted $\operatorname{Sym}(n)$. Thus, the horizontal bundle is

$$\operatorname{Hor}_A = \{\dot{A} = UA \in T_A \operatorname{GL}(n) : U \in \operatorname{Sym}(n)\}.$$

Note that the horizontal bundle is independent of Σ_0: the dependence of the metric on Σ_0 is compensated by the dependence of the projection π on Σ_0.

The *polar cone* $K_\Diamond$ consists of all matrices in $\operatorname{GL}(n)$ connected with the identity by a horizontal geodesic, i.e., geodesics $\gamma(t)$ with $\dot{\gamma}(0) \in \operatorname{Hor}_I$.

The following lemma (see Modin [11] for a geometric proof) is vital to understand the importance of $K_\Diamond$.

Lemma 2.2. *The restriction of the projection π to $K_\Diamond$ is an isomorphism. In other words, $K_\Diamond$ is a section of the principal bundle (6).*

In fact, the polar cone *itself* consists of positive definite symmetric matrices (geometrically these elements are not, however, elements of the base space). Now, let $A \in \operatorname{GL}(n)$ be arbitrary and set $\pi(A) = \Sigma_1$. By Lemma 2.2, there is a unique matrix $K_\Diamond \ni P = \pi|_{K_\Diamond}^{-1}(\Sigma_1)$. It is clear that P and A both are in the same fiber. Therefore, $Q = P^{-1}A \in O(n, \Sigma_0)$, by the isomorphism (5). Thus, as $PQ = A$, we have obtained the polar decomposition. P lies at the intersection of the polar cone and the fiber $\pi^{-1}(A)$. The uniqueness of the polar decomposition means, geometrically, that each fiber intersects the polar cone only once.

3. **Vertical gradient flow.** Having introduced the geometry of the Gaussian Monge problem, we move to the gradient flow. Let $\nabla_\mathcal{G}$ denote the gradient with respect to the Riemannian metric (4). Further, we denote by $\operatorname{Proj}_{\mathrm{Ver}}$ the bundle map on $T\operatorname{GL}(n)$ given by orthogonal projection onto the vertical distribution.

The following vertical gradient flow is suggested in Modin [11]:

$$\dot{B} = -\operatorname{Proj}_{\mathrm{Ver}} \nabla_\mathcal{G} J(B), \;\; B(0) = A. \tag{7}$$

Note that $\operatorname{Proj}_{\mathrm{Ver}} \nabla_\mathcal{G} J(B)$ is exactly the gradient for the restriction of the metric $\mathcal{G}$ and the function $J(B)$ to the fiber of A. Therefore, the projected gradient flow (7) is itself a gradient flow on the Riemannian sub-manifold given by the fiber.

Our goal is to study Equation (7). The main result is the following.

Theorem 3.1. *Let A be in the identity component of $\operatorname{GL}(n)$. Then the gradient flow (7) has the following properties:*

1. *A global solution exists.*
2. *It converges as $t \to \infty$ to the matrix P in the polar decomposition of A.*

For an illustration of the vertical gradient flow, as well as the geometry of the Gaussian Monge problem, see Figure 1.

3.1. **Analysis of the gradient flow.** To prove Theorem 3.1, we leverage the geometry described in Section 2.1. The first step is to explicitly determine the gradient flow in Equation (7). To this end, we take the time derivative of $J(B)$, defined in Equation (2),

$$\frac{\mathrm{d}}{\mathrm{d}t} J(B) = 2\operatorname{Tr}(\Sigma_0 (B - I)^T \dot{B}) = \mathcal{G}_B(2(B - I), \dot{B}).$$

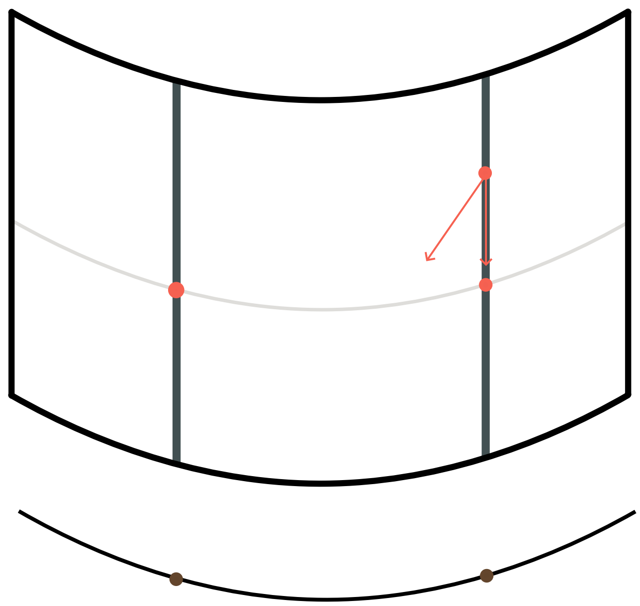

FIGURE 1. Illustration of the geometry of the problem. Note how the polar cone intersects the fiber at one point, corresponding to the uniqueness of the polar factorization.

Thus, the vertical gradient flow is

$$\dot{B} = -2(B - I) - SB, \ B(0) = A, \tag{8}$$

where S is a Lagrange multiplier forcing B to remain on the fiber.

It is advantageous to rewrite the gradient flow in terms of a right-reduced variable, both from a theoretical and computational perspective. Therefore, we introduce the variable $\Omega := \dot{B}B^{-1}$ which, by the definition of the vertical bundle, is in $\mathfrak{o}(n, \Sigma_1)$. Algebraic manipulations of Equation (8), see Modin [11, Section 2.4.1] for details, yield the equivalent right–reduced gradient flow,

$$\dot{B} = \Omega B, \ B(0) = A, \tag{9}$$

$$\Sigma_1\Omega + \Omega\Sigma_1 = 2\Sigma_1(B^{-1} - B^{-T}). \tag{10}$$

Finally, note that the right-hand side of Equation (10) may be expressed without matrix inverses of B. Indeed, it holds that

$$B^{-1} = \Sigma_0 B^T A^{-T} \Sigma_0^{-1} A^{-1}. \tag{11}$$

We are thus in position to give the proof.

Proof of Theorem 3.1. First we prove global existence. We claim that the vector field in Equation (9) is locally Lipschitz continuous. To see this, note that it is given by the product of Ω, a function of B as defined by Equation (10), and B. If we can verify that $B \mapsto \Omega$ is a Lipschitz map, then the claim follows, as the map $B \mapsto B$ is trivially Lipschitz and the product of two Lipschitz maps is locally Lipschitz.

The Sylvester equation (10) is equivalent to the linear system

$$\mathcal{S}\,\mathrm{Vec}(\Omega) = \mathrm{Vec}(2\Sigma_1(B^{-1} - B^{-T})), \tag{12}$$

where $\mathcal{S} = I \otimes \Sigma_1 + \Sigma_1 \otimes I$ and $\mathrm{Vec} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n^2}$ denotes the vectorization of a matrix. Here $\otimes$ is the Kronecker product of matrices [8, Chapter 4.3]. The linear system (12) has a unique solution [3], so $\Omega = \mathrm{Vec}^{-1}(\mathcal{S}^{-1}\,\mathrm{Vec}(2\Sigma_1(B^{-1} - B^{-T})))$. The map $\mathbb{R}^{n^2} \ni x \mapsto \mathcal{S}^{-1}x$ as well as vectorization and its inverse are Lipschitz maps. Finally, $B \mapsto 2\Sigma_1(B^{-1} - B^{-T}))$ is Lipschitz since B^{-1} can be computed from B and the fixed matrices Σ_0 and A using Equation (11). Thus, $B \mapsto \Omega$ is a Lipschitz map.

Recall that the gradient flow is on the fiber $\pi^{-1}(\Sigma_1)$ which is isomorphic to the Lie group $\mathrm{O}(n, \Sigma_0)$, on which the metric (4) is right-invariant. On a Lie group with a right-invariant metric, gradient flows that are locally Lipschitz with respect to the topology induced by the metric admit global existence [2, Thm. 2.7].

To prove convergence, consider the functional $J(B(t))$ evaluated along the flow as a function of time. A general fact of gradient flows is that the functional is decreasing in time, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} J(B) \leq 0.$$

We know that $J(B) = d^2(B, I)$ is bounded from below; as $B \in \pi^{-1}(\Sigma_1)$, $J(B)$ must be larger than or equal to $D_{\Sigma_1} := d^2(\pi^{-1}(\Sigma_1), \pi^{-1}(I))$, the squared distance between $\pi^{-1}(\Sigma_1)$ and the identity fiber. Therefore, $J(B)$ converges as $t \to \infty$. The idea is now to show that $\frac{\mathrm{d}}{\mathrm{d}t} J(B)$ converges to 0. By the chain rule one sees that

$$\frac{\mathrm{d}}{\mathrm{d}t} J(B) = 2\,\mathrm{Tr}(\Sigma_0 (B - I)^T \Omega B) = 2\,\mathrm{Tr}(\Sigma_1 \Omega) - 2\,\mathrm{Tr}(\Sigma_0 \Omega B) = -2\,\mathrm{Tr}(\Sigma_0 \Omega B), \tag{13}$$

where the second-to-last inequality follows by noting that $\pi(B) = \Sigma_1$ and the final since $\Omega\Sigma_1$ is skew-symmetric and thus has zero trace. The second derivative of J is

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} J(B) = -2\,\mathrm{Tr}(\Sigma_0 \dot{\Omega} B) - 2\,\mathrm{Tr}(\Sigma_0 \Omega^2 B),$$

where $\dot{\Omega}$ is obtained by solving the Sylvester equation arising from the time derivative of Equation (10),

$$\Sigma_1 \dot{\Omega} + \dot{\Omega} \Sigma_1 = 2\Sigma_1 \frac{\mathrm{d}}{\mathrm{d}t}(B^{-1} - B^{-T}).$$

Note that the Frobenius norm of a solution to a Sylvester equation can be bounded by the Frobenius norm of the right-hand side. Indeed, if we let

$$\mathrm{Sep}(\Sigma_1) := \min_{X \in \mathbb{R}^{n\times n}} \|X\Sigma_1 + \Sigma_1 X\|_F / \|X\|_F,$$

then

$$\|\dot{\Omega}\|_F \leq \frac{\|2\Sigma_1 \frac{\mathrm{d}}{\mathrm{d}t}(B^{-1} - B^{-T})\|_F}{\mathrm{Sep}(\Sigma_1)}, \tag{14}$$

$$\|\Omega\|_F \leq \frac{\|2\Sigma_1 (B^{-1} - B^{-T})\|_F}{\mathrm{Sep}(\Sigma_1)}. \tag{15}$$

In our case, $\mathrm{Sep}(\Sigma_1) = 2\lambda_{\min}(\Sigma_1)$ [12], where $\lambda_{\min}(\Sigma_1)$ denotes the smallest eigenvalue of Σ_1. Now,

$$\left| \frac{\mathrm{d}^2}{\mathrm{d}t^2} J(B) \right| \leq 2|\,\mathrm{Tr}(\Sigma_0 \dot{\Omega} B)| + 2|\,\mathrm{Tr}(\Sigma_0 \Omega^2 B)| \leq 2\|\Sigma_0\|_F \left(\|\dot{\Omega}\|_F + \|\Omega\|_F^2 \right) \|B\|_F$$

where the final inequality is due to the Cauchy–Schwarz inequality and the fact that the Frobenius norm is sub–multiplicative. Further, note that Equation (14) and Equation (15) imply that

$$\|\Omega\|_F \le \frac{\|\Sigma_1\|_F}{\lambda_{\min(\Sigma_1)}}\|B^{-1} - B^{-T}\|_F$$

and

$$\|\dot{\Omega}\|_F \le \frac{\|\Sigma_1\|_F}{\lambda_{\min(\Sigma_1)}}\left\|\frac{\mathrm{d}}{\mathrm{d}t}(B^{-1} - B^{-T})\right\|_F$$

Recalling Equation (11), it follows by the triangle inequality and the trace-invariance of the Frobenius norm that

$$\|B^{-1} - B^{-T}\|_F \le 2\|B^{-1}\|_F = 2\|\Sigma_0 B^T A^{-T}\Sigma_0^{-1}A^{-1}\|_F$$

so again by applying that the Frobenius norm is sub–multiplicative, one arrives at

$$\|\Omega\|_F \le C_{\Sigma_1,\Sigma_0,A}\|B\|_F,$$

where $C_{\Sigma_1,\Sigma_0,A} = 2\|\Sigma_1\|_F\|\Sigma_0\|_F\|\Sigma_0^{-1}\|_F\|A^{-1}\|_F^2/\lambda_{\min}(\Sigma_1)$. Similarly, Equation (9) to replace $\dot{B}$ with ΩB together with Equation (11) gives that

$$\|\dot{\Omega}\|_F \le C^2_{\Sigma_1,\Sigma_0,A}\|B\|_F^2,$$

Therefore,

$$\left|\frac{\mathrm{d}^2}{\mathrm{d}t^2}J(B)\right| \le 2C^2_{\Sigma_1,\Sigma_0,A}\|\Sigma_0\|_F\|B\|_F^3,$$

which is finite as $\|B\|_F \le C\,\mathrm{Tr}(\Sigma_0 B^T B) = C\,\mathrm{Tr}(\Sigma_1) < \infty$ for some constant $C > 0$. As the second derivative is bounded, Barbalat's lemma [10, Lemma 8.2] gives that $\lim_{t\to\infty}\frac{\mathrm{d}}{\mathrm{d}t}J(B) = 0$. However, the time derivative being zero at some $B \in \pi^{-1}(\Sigma_1)$ means, by Equation (13), that

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}J(B) = \mathrm{Tr}(\Sigma_0\Omega B) = \mathrm{Tr}(\Omega\Sigma_1 B^{-T}).$$

where the second inequality is due to the identity $B\Sigma_0 = \Sigma_1 B^{-T}$ and the invariance of the trace under cyclic permutations. As $\Omega\Sigma_1$ by construction is skew-symmetric, B^{-T} must be symmetric, i.e., B is symmetric as well as in the fiber $\pi^{-1}(\Sigma_1)$. From Theorem 2.1 it follows that there is only one such matrix, namely P. Geometrically, it means that P is the unique point of intersection between the fiber and the polar cone. Thus, the flow $B(t)$ approaches the matrix P. □

4. **A numerical experiment.** In this section, we provide a numerical example of the gradient flow, illustrating its convergence. For convenience, let $\Sigma_0 = I$. To avoid selecting a biased initial matrix A, we instead draw several random matrices with normally distributed elements. The dimension of each matrix is randomly chosen among 2, 4, 8 and 16. The matrix generation process is continued until a set $\{A^i, i = 1,\ldots,1000\}$ of 1000 invertible matrices of different sizes is obtained.

Equation (9) is discretized by the *Lie–Euler method* [9],

$$B^i_{k+1} = \exp(h\Omega^i_k)B^i_k, \quad B^i_0 = A^i.$$

Here, exp refers to the matrix exponential, and we set $h = 0.05$ and integrate the flow for 600 steps. Further, Ω_k, given as the solution to Equation (10) is computed using the Bartels–Stewart algorithm [3]. We compute the polar decomposition P^i

of each A^i from its SVD factorization, as suggested in [7, Section 3.1]. We thus obtain approximated flow curves B_k^i for $k = 1, \dots, 600$, for each $i = 1, \dots, 1000$.

In Figure 2, the squared distance between each B_k^i and P^i is plotted against time $t_k = hk$, as well as the median, and upper and lower 10% percentiles over all curves in each time step. This illustrates the convergence of the gradient flow.

Figure 2 indicates exponential convergence. Further, it indicates that the exponential convergence rate depends on the dimension of the matrix, as the lines cluster in some distinct bands. A full analysis of these issues is beyond the scope of this paper.

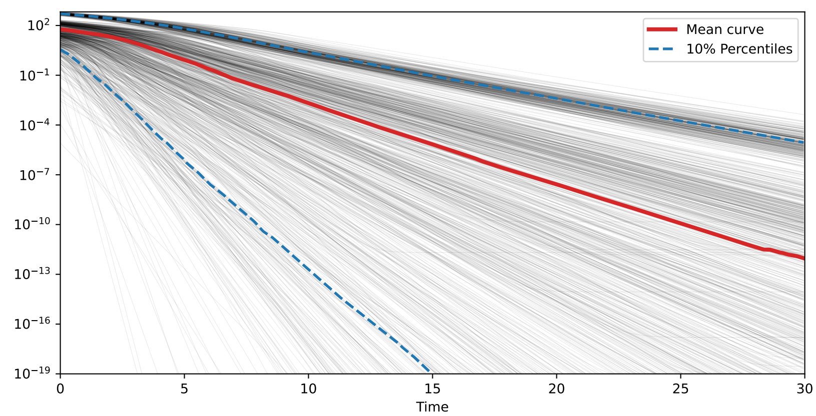

FIGURE 2. Plot illustrating the convergence of 1000 gradient flow paths, each starting in a randomly selected matrix. The squared distance between each path and its corresponding polar decomposition is plotted (black curves) as well as the median squared distance (red curve) and percentiles (blue curves).

5. **Conclusion and outlook.** In the present paper, we have examined a gradient flow and proved that it "computes" the polar decomposition of matrices. It should be noted that this way to compute the polar decomposition is inefficient compared to existing numerical linear algebra based algorithms. Instead, the main motivations for the study are the following. First, it shows how the polar decompositions is directly tied to the Gaussian Monge problem via its dynamical formulation. Second, what we have considered is a finite-dimensional restriction of the full, infinite-dimensional optimal transport problem in the smooth category, for which the same geometric structure holds, see Modin [11]. This paper may thus be seen as a pre-study for a detailed analysis of the vertical gradient flow in the infinite-dimensional setting, which could give rise to new algorithms or results for optimal mass transport.

REFERENCES

[1] S. Amari, *Information Geometry and Its Applications*, Springer, 2016.
[2] T. Balehowsky, C-J. Karlsson, K. Modin, Shape analysis via gradient flows on diffeomorphism groups, *Nonlinearity*, **36** (2023), 862-877.

[3] R. H. Bartels and G. W. Stewart, Solution of the matrix equation $AX + XB = C$, *Commun. ACM*, **15** (1972), 820-826.

[4] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, *Comm. Pure Appl. Math.*, **44** (1991), 375-417.

[5] R. W. Brockett, Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems, *Linear Algebra Appl.*, **146** (1991), 79-91.

[6] H. Flaschka, The Toda lattice. I. Existence of integrals, *Phys. Rev. B*, **9** (1974), 1924-1925.

[7] N. J. Higham, Computing the polar decomposition– with applications, *SIAM J. Sci. and Stat. Comput*, **7** (1986), 1160-1174.

[8] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis,* Cambridge University Press, Cambridge, 1991.

[9] A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett and A. Zanna, Lie-group methods, *Acta Numer.*, **9** (2000), 215-365.

[10] H. K. Khalil, *Nonlinear Systems,* 3rd edition, Pearson, 2001.

[11] K. Modin, Geometry of matrix decompositions seen through optimal transport and information geometry, *J. Geom. Mech.*, **9** (2017), 335-390.

[12] J. M. Varah, On the separation of two matrices, *SIAM J. Numer. Anal.*, **16** (1979), 216-222.

Received June 2023; revised October 2023; early access November 2023.