Geometry-preserving lie group integrators for differential equations on the manifold of symmetric positive definite matrices

Lucas Drumetz, Alexandre Reiffers-Masson, Naoufal El Bekri, Franck Vermet

GEOMETRY-PRESERVING LIE GROUP INTEGRATORS FOR DIFFERENTIAL EQUATIONS ON THE MANIFOLD OF SYMMETRIC POSITIVE DEFINITE MATRICES

Lucas Drumetz[1], *Alexandre Reiffers-Masson*[1], *Naoufal El Bekri*[2], *Franck Vermet*[2]

[1]IMT Atlantique, Lab-STICC, UMR CNRS 6285, Brest, France
[2]Univ Brest, UMR CNRS 6205, Laboratoire de Mathématiques de Bretagne Atlantique, France

ABSTRACT

In many applications, one encounters signals that lie on manifolds rather than a Euclidean space. In particular, covariance matrices are examples of ubiquitous mathematical objects that have a non Euclidean structure. The application of Euclidean methods to integrate differential equations lying on such objects does not respect the geometry of the manifold, which can cause many numerical issues. In this paper, we propose to use Lie group methods to define geometry-preserving numerical integration schemes on the manifold of symmetric positive definite matrices. These can be applied to a number of differential equations on covariance matrices of practical interest. We show that they are more stable and robust than other classical or naive integration schemes on an example.

Index Terms— Lie groups, differential equations, symmetric positive definite matrices, stochastic differential equations

1. INTRODUCTION

Ordinary Differential Equations (ODEs) arise everywhere in science and signal processing, and are a fundamental tool to describe the evolution of dynamical systems [1]. However, except in simple cases, they rarely have analytical solutions, and numerical integration schemes are required to obtain approximate solutions. Most of the time, the variable to integrate lives in a Euclidean space, typically $\mathbb{R}^n$. In this situation, one can choose from many schemes that have been developed over the years, ranging from simple explicit/implicit Euler or Runge-Kutta methods to adaptive time step schemes [2].

However, in a certain number of cases of practical interest, one may require that the variable to integrate lies on a manifold [3]. Examples include, among others, flows on spheres, rotation or covariance matrices (or other matrix manifolds)[4]. For abstract manifolds (i.e. that are not seen as embedded submanifolds of $\mathbb{R}^n$), one cannot use classical integration schemes, however advanced, since those require a vector space structure to compute additions and scalar multiplications. In the more common and intuitive case of embedded submanifolds of $\mathbb{R}^n$, the underlying vector space structure makes it possible to apply such ODE integration methods, but they cannot guarantee that the numerical solution remains on the manifold at each time step. For example, considering a differential equation on a sphere, a small step taken in a direction of a tangent vector to the current point does not belong to the sphere anymore. A small error may be tolerable in practice, but staying on the manifold may be crucial for subsequent uses of the solution, e.g. computing geodesic distances for Riemannian manifolds [5], for asymptotic stability [6], or simply keeping the structural or physical interpretation of a variable.

More formally, one may be interested in obtaining the flow of a smooth vector field on a smooth manifold $\mathcal{M}$ generated by a differential equation, with the initial condition $x(0) = x_0$:

$$\frac{dx}{dt} = F|_{x(t)}(t), \tag{1}$$

where $x \in \mathcal{M}$ and $F|_{x(t)}(t)$ is a (possibly time dependent) tangent vector to the manifold at x, and we have $F : [0, +\infty[\to \mathfrak{X}(\mathcal{M})$, with $\mathfrak{X}(\mathcal{M})$ the set of smooth vector fields on $\mathcal{M}$ [7].

To integrate such differential equations, several frameworks were developed under the umbrella term of Lie group integrators [8]. Interestingly, most of these methods can be extended to any smooth manifold acted upon by a transitive Lie group [4]. We will use this latter framework to avoid having to define a Lie group structure on our manifold of interest, which can be done [9], but is less flexible than what we propose here.

In this paper, we focus more specifically on the manifold of $n \times n$ symmetric positive definite (SPD) matrices, denoted as Sym_n^+ [10]. This manifold is of particular interest since it is the manifold of (nondegenerate) covariance matrices, that are fundamental for multivariate statistics. Flows of covariance matrices arise in many different applications, such as Brain Computer Interfaces (BCI) [11], Diffusion Tensor Image processing [12], finance [13], control [14], or data assimilation [15], to represent the evolution of second order moments of random variables. For example, one may be interested in the second order moments of the solution of Stochastic Differential Equations (SDEs), because they provide simplified and interpretable representations of stochastic processes, though partial in general. In data assimilation, quantifying and propagating the uncertainty of the variable to reconstruct, coming from either the dynamical prior model or the observations is crucial and is done in practice using covariance matrices [16].

This work was supported by Agence Nationale de la Recherche under grant ANR-21-CE48-0005 LEMONADE.

Solutions of covariance matrix ODEs which are not SPD are meaningless in terms on statistical interpretation, strengthening the need for algorithms with guarantees. Thus, we focus on equations similar to (1), where the manifold $\mathcal{M}$ is Sym_n^+, and the right hand side of (1) is a symmetric matrix (an element of the tangent space of the SPD manifold at the current point).

Our contributions are threefold: i) We propose to use a Lie group action of invertible matrices on Sym_n^+ as a basis, as it is broadly applicable to many equations of interest. ii) From there, we design a Lie group version of the Runge-Kutta 4 (RK4) method (applicable to many other schemes) on Sym_n^+. iii) We conduct experiments an example ODE on Sym_n^+ related to a multivariate SDE. They indicate that our integrators perform better than classical schemes, in particular when the integration step is large. In spite of the ubiquiteness of covariance matrices, to the best of our knowledge, Lie group integrators have not been considered yet for Sym_n^+. A reason might be that for small enough time steps, iterates of classical methods remain in the manifold. We have even obtained a sufficient condition on the step size to guarantee this for an explicit Euler method. However, with moderately big time steps, classical methods may cross the boundary of Sym_n^+, (consisting of symmetric semipositive definite matrices), leading to low quality solutions or even diverging algorithms.

2. BACKGROUND ON LIE GROUP INTEGRATORS

The general idea behind Lie group methods is to take advantage of the fact that the flow of a simple class of vector fields on the Lie group is easy to compute via the Lie exponential map. The corresponding equations are analogous to the linear ODE $d\mathbf{x}/dt = \mathbf{A}\mathbf{x}$ in Euclidean spaces. In the case of general ODE, discretizing by temporarily fixing the vector field of a general equation with a nonconstant "$\mathbf{A}$" (depending on x and t), an approximate Euler-like scheme can be computed step by step. Interestingly, these methods can be effortlessly extended to smooth manifolds on which we can find a *transitive Lie group action*, without more structure on the manifold [7, 8]. We use this property to define integrators on the manifold of SPD matrices. Throughout this section, we follow [4] (Chap.2). Let $\mathcal{M}$ be the smooth manifold, G the Lie group, and I the neutral element of G. A map $\Lambda: G \times \mathcal{M} \to \mathcal{M}$ is a Lie group action if and only if it is smooth and satisfies the two conditions:

$$\forall\, x \in G, \Lambda(I, x) = x, \tag{2}$$

$$\forall\, x, y, z \in G, \Lambda(y, \Lambda(z, x)) = \Lambda(yz, x). \tag{3}$$

Λ is said to be transitive if

$$\forall\, x, y \in \mathcal{M}, \exists g \in G, \Lambda(g, x) = y. \tag{4}$$

This means that any point of the manifold can be reached from any other using the group action with an element of the group.

To every Lie group G is associated a Lie algebra $\mathfrak{g}$, which is a vector space (the tangent space to the Lie group at the identity) that represents an infinitesimal vector description of the group. Consequently, associated to a transitive Lie group action on a smooth manifold is a *Lie algebra homomorphism* that translates the group action into an infinitesimal one from the Lie algebra, giving an element of tangent space to the manifold at every point. It determines, from the group action, the type of equations that can be dealt with. It is defined [4] (Lemma 2.6) as a map $\lambda_*: \mathfrak{g} \to \mathfrak{X}(\mathcal{M})$ such that, for a given point $x \in \mathcal{M}$:

$$\lambda_*(a)(x) = \frac{d}{ds}\Lambda(\rho(s), x)|_{s=0}, \tag{5}$$

where $\rho(s)$ is a smooth curve on G, parameterized by a scalar s, with initial value $\rho(0) = I$ and initial speed $a \in \mathfrak{g}$ ($\rho'(0) = a$). Intuitively, this curve represents a direction a on the Lie algebra towards which we can move infinitesimally from any point in G. λ_* can be seen as an infinitesimal group action that extends this idea of moving infinitesimally from the current point x on the manifold $\mathcal{M}$ in a specific way.

The general procedure to build a Lie group scheme to compute x_{i+1} from x_i, with $t_{i+1} = t_i + \Delta t$, with Δt being the time step is:

1) Write the differential equation in terms of a Lie algebra homomorphism with an adequately chosen group action (Λ, with the associated λ_*)

$$\frac{dx}{dt} = F|_{x(t)}(t) = \lambda_*(\xi(x(t), t))(x(t)), \tag{6}$$

with the initial condition $x(t_i) = x_i$, and where $\xi: \mathcal{M} \to \mathfrak{g}$ is a smooth function (possibly time dependent). At a given step, let us temporarily fix ξ to its current value $\xi(x_i, t_i)$.

2) Thanks to [4] (Lemma 2.7), we can show there exists a curve $\gamma \in G$ starting at the identity, with initial speed $\xi(x_i, t_i)$, such that we can write the flow on $\mathcal{M}$ induced by $\xi(x_i, t_i)$ as $x(t_i + \Delta t) = \Lambda(\gamma(\Delta t), x_i)$. In addition, $\gamma(t) \in G$ follows a specific differential equation. To write its expression, we need to define a "product" between an element a of the Lie algebra $\mathfrak{g}$ and an element σ of the Lie group G:

$$a\sigma = \frac{d}{ds}\rho(s)\sigma|_{s=0}, \tag{7}$$

with ρ a curve in G initially at I and with initial speed a. Then, the differential equation followed by γ (with $\gamma(0) = I$) writes:

$$\frac{d\gamma}{dt} = \xi(x_i, t_i)\gamma(t). \tag{8}$$

3) The solution of the ODE (8) on G is [4] (Theorem 2.8):

$$\gamma(\Delta t) = \exp(\Delta t \xi(x_i, t_i)), \tag{9}$$

with exp the Lie group exponential map. Thus, computing the flow of this ODE on G is easy, provided the exponential map is tractable. Of course, the solution will be an accurate

approximation of the flow of Eq. (1) if Δt is sufficiently small, since throughout the procedure $\xi(x_i, t_i)$ is kept constant.
4) We finally come back to the manifold using the group action:

$$x_{i+1} = x(t_i + \Delta_t) = \Lambda(\gamma(\Delta_t), x_i). \tag{10}$$

In practice: We simply need to compute (9) for the current value of $\xi(x_i, t_i)$, and come back to the manifold via (10). Then, ξ is updated to repeat the procedure. For other explicit schemes requiring intermediary values the flow, e.g. RK4, we can compute them as above, using the right values of x and t.

3. APPLICATION TO THE SPD MANIFOLD

Here, we propose and examine a suitable Lie group action on Sym_n^+ to build Lie group integration schemes on Sym_n^+. First, the set of SPD matrices is indeed a smooth manifold, whose tangent space at each point can be identified with the set of symmetric matrices Sym_n. Thus, any differential equation on Sym_n^+ has a symmetric matrix as a right-hand side.
We choose the Lie group acting on the manifold to be the set of invertible matrices, endowed with the usual matrix multiplication (i.e. the general linear group $GL_n(\mathbb{R})$). Its Lie algebra is the set of all square matrices $\mathcal{M}_n(\mathbb{R})$, which we will simply denote as $\mathbb{R}^{n\times n}$. We define our group action:

$$\begin{aligned} \mathbf{\Lambda} :\ GL_n(\mathbb{R}) \times \mathrm{Sym}_n^+ &\to\ \mathrm{Sym}_n^+ \\ (\mathbf{M}, \mathbf{P}) &\mapsto \mathbf{M}\mathbf{P}\mathbf{M}^T. \end{aligned} \tag{11}$$

We can easily check that it satisfies all the requirements to be a proper group action and that it is indeed transitive. This group action is very natural for covariance matrices, as it corresponds to the effect of an invertible linear transformation of a random vector on its covariance matrix.
From this transitive group action, using the definition in (5), we obtain the Lie algebra homomorphism $\boldsymbol{\lambda}_* : \mathbb{R}^{n\times n} \to \mathfrak{X}(\mathrm{Sym}_n^+)$ (applied at the tangent space of a point $\mathbf{P} \in \mathrm{Sym}_n^+$):

$$\boldsymbol{\lambda}_*(\mathbf{M})(\mathbf{P}) = \frac{d}{ds}\mathbf{\Lambda}(\boldsymbol{\gamma}(s), \mathbf{P})|_{s=0} = \mathbf{M}\mathbf{P} + \mathbf{P}\mathbf{M}^T, \tag{12}$$

where $\boldsymbol{\gamma}(s) = \mathbf{I} + s\mathbf{M} + ...$ is a smooth curve on the Lie group starting form the identity and with initial speed $\mathbf{M} \in \mathbb{R}^{n\times n}$. Since we are dealing with a matrix group, embedded in $\mathbb{R}^{n\times n}$, we can write the curve as a Taylor expansion. Following Eq. (6), we can tackle any equation of the form

$$\frac{d\mathbf{P}}{dt} = \boldsymbol{\xi}(\mathbf{P}, t)\mathbf{P} + \mathbf{P}\boldsymbol{\xi}(\mathbf{P}, t)^T, \tag{13}$$

with $\boldsymbol{\xi} : \mathrm{Sym}_n^+ \to \mathbb{R}^{n\times n}$ any smooth function (possibly time dependent) from SPD matrices to $\mathbb{R}^{n\times n}$. This class of functions is quite broad, so Eq. (13) is not very restrictive and many equations of interest can be written this way. For instance, with a constant $\boldsymbol{\xi}$, Eq. (13) governs the dynamics of the covariance of a random variable that propagates via a deterministic linear dynamical system [16]. More complex functions $\boldsymbol{\xi}$ can model more complex situations. From (8), we consider an initial condition $\mathbf{P}(t_i) = \mathbf{P}_i$. The differential equation on $\boldsymbol{\gamma}$ is

$$\frac{d\boldsymbol{\gamma}}{dt} = \boldsymbol{\xi}(\mathbf{P}_i, t_i)\boldsymbol{\gamma}(t), \tag{14}$$

with $\boldsymbol{\gamma}(0) = \mathbf{I}$, and the product (7) reduces to the matrix product, since we consider a matrix Lie group.
Finally, following (9), the solution $\boldsymbol{\gamma}$ is, as one would expect,

$$\boldsymbol{\gamma}(\Delta t) = \exp(\Delta t\boldsymbol{\xi}(\mathbf{P}_i, t_i)), \tag{15}$$

where here, $\exp$ is simply the matrix exponential. Finally for an explicit Euler scheme, we apply step 4 of Sec. 2 to obtain the next iterate on Sym_n^+:

$$\mathbf{P}_{i+1} = \mathbf{P}(t_i + \Delta t) = \mathbf{\Lambda}(\boldsymbol{\gamma}(\Delta t), \mathbf{P}_i). \tag{16}$$

For RK4, we compute each of the required intermediary values using Eqs. (15) and (16) updating the values of $\mathbf{P}$ and t.

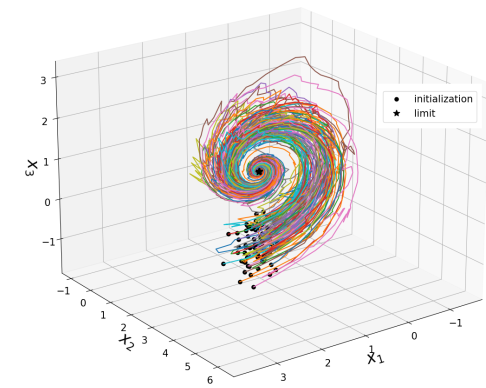

Fig. 1. 100 trajectories of the SDE (17) with our choice of parameters and a small time step (10 times larger than in the next figures).

4. CASE STUDY

We are interested here in a multivariate generalization of a Geometric Brownian Motion (GBM), given by the (Itô) SDE [17]:

$$d\boldsymbol{X} = \left(\mathbf{A} + \frac{1}{2}\mathbf{B}^2\right)\boldsymbol{X}dt + \mathbf{B}\boldsymbol{X}dW_t, \tag{17}$$

with $\boldsymbol{X}$ a random vector of size n, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ two commuting matrices, such that the eigenvalues of $\mathbf{A} + \frac{1}{2}\mathbf{B}^2$ have a strictly negative real part. $W_t \in \mathbb{R}$ is a Brownian motion. A closed form expression of the trajectories exists, see Fig. 1.

We can derive ODEs followed by the mean (taking expectations in (17)) and covariance matrix (using Itô's Lemma [18] on $\boldsymbol{X}\boldsymbol{X}^T$, taking expectations, and a few algebraic manipulations). They provide a broad summary of the statistics of the process (much simpler than e.g. the Fokker-Planck equation):

$$\frac{d\mathbf{m}}{dt} = \left(\mathbf{A} + \frac{1}{2}\mathbf{B}^2\right)\mathbf{m} \tag{18}$$

$$\frac{d\mathbf{P}}{dt} = \left(\mathbf{A} + \frac{1}{2}\mathbf{B}^2\right)\mathbf{P} + \mathbf{P}\left(\mathbf{A} + \frac{1}{2}\mathbf{B}^2\right)^T + \mathbf{B}(\mathbf{P} + \mathbf{m}\mathbf{m}^T)\mathbf{B}^T \tag{19}$$

Eq. (19) can indeed be put in the form of Eq. (13), using

$$\boldsymbol{\xi}(\mathbf{P}) = \left(\mathbf{A} + \frac{1}{2}\mathbf{B}^2\right) + \frac{1}{2}\mathbf{B}(\mathbf{P} + \mathbf{m}\mathbf{m}^T)\mathbf{B}^T\mathbf{P}^{-1}. \tag{20}$$

Due to space constraints, we can only detail this example (with $n = 3$) here, but our method also applies e.g. to model the covariance of a multivariate Ornstein-Uhlenbeck process [19], or to several types of Riccati equations encountered in control [14], with appropriate choices of $\boldsymbol{\xi}$.

We consider three explicit RK4 schemes for Eq. (19): a Euclidean scheme, a variant where each step is brought back to the manifold using a Riemannian exponential map, with the affine invariant metric of [10, 20] (Riemmannian RK4), and our method (Lie RK4). We compare them to a reference numerical solution obtained from a classical RK4 method with a very small time step, for which we know the trajectory remains on Sym_n^+ and the integration error is small. We choose specific commuting $\mathbf{A}$ and $\mathbf{B}$ so all the eigenvalues of $\mathbf{A} + \frac{1}{2}\mathbf{B}^2$ have a strictly negative real part. We start from a Gaussian initial condition $\boldsymbol{X}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$. Then, the process converges to a distribution given by a Dirac centered at $\mathbf{0}$.

We set $t \in [0, 5]$, and show results with 30 evenly spaced time steps. We show the trajectories of the covariance matrix for the three competing methods in Fig. 2, and plot three different distances between the trajectories and the reference in Fig. 3: the Frobenius (Euclidean) distance in $\mathbb{R}^{n\times n}$, as well as two Riemannian distances on Sym_n^+: the log Euclidean [9] and the affine invariant [10] distances, that better account for the geometry of the trajectories (and give similar values in this example, but not for larger time steps). At first glance, from Fig. 2, all methods seem to reasonably approximate the reference, with a notably worse performance for the Riemannian RK4. In terms of Euclidean distance (Fig. 3 (a)), the classical RK4 is only slightly worse than the proposed method. However, looking at Fig. 3 (b) and (c), we see that Riemannian-RK4 is actually much worse than Lie RK4, and the classical RK4 leaves the manifold after a few iterations only (at least an eigenvalue becomes negative), resulting in infinite distances. Even though the trajectory is not far from the true one in Euclidean distance, the solution is not a covariance anymore and loses any statistical interpretation. Errors keep increasing since for the Riemannian distances, the stationary zero matrix is outside the manifold, at infinite distance from it.

With even larger time steps, both the Euclidean and the Riemannian-RK4 diverge after a small number of iterations. For the former, this is expected since the manifold structure

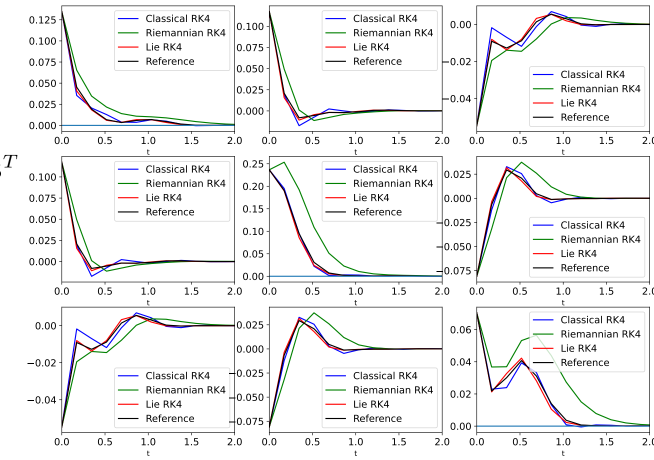

Fig. 2. Entries (i, j) of the covariance matrix across time for all schemes and the reference. On diagonal terms, the blue line cannot be crossed (this would result in a negative variance).

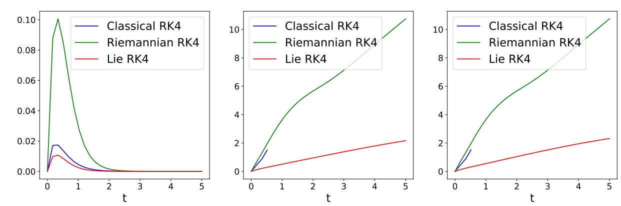

Fig. 3. (a) Frobenius (b) Log Euclidean (c) Affine Invariant distances between each integrated trajectory and the reference.

is destroyed even earlier. The latter fails because it is not truly intrinsic to the manifold, contrary to ours, so a very bad step in the Euclidean domain cannot be made up for, and the (Riemannian) exponential map may not be optimal to provide accurate schemes. For smaller time steps, Lie RK4 remains the most precise up to a point, then classical RK4 becomes slightly better on all metrics, probably due to accumulating errors during the additional computations. However, in many applications, the step size is imposed by the problem, e.g. when observation data have a low sampling rate.

5. CONCLUSION

We have presented a Lie group framework to define structure-preserving integration schemes for flows of SPD matrices. Our fully intrinsic method keeps iterates on the manifold, and provides smaller integration error than classical or naive methods, especially for large time steps. This will be useful in our future work to learn and represent uncertainty in data assimilation [21, 22] or controls from observation data when governing equations are unknown (by learning a function $\boldsymbol{\xi}$ matching the data). In such cases, the time step is imposed by data and the training process may lead to ill-conditioned equations.

6. APPENDICES

6.1. Sufficient condition on the integration step for the Euclidean explicit Euler method to stay on Sym_n^+

In this section, we prove a theorem that provides a sufficient condition on the integration step ρ for the classical explicit Euler method to yield an iterate that still belongs to Sym_n^+, regardless of the equation under consideration . $\mathbf{T}$ has to be understood to be the right hand side of the ODE, computed for the current value of the matrix to be integrated. Tighter bounds could probably be found for a specific equation by looking at the exact expression of $\mathbf{T}$.

Theorem 6.1. *Let $\mathbf{P} \in \mathrm{Sym}_n^+$, and $\mathbf{T} \in \mathrm{Sym}_n$, i.e. $\mathbf{T}$ is in the tangent space of $\mathbf{P}$. Then*

- *if $\mathbf{T}$ is positive semidefinite, then $\forall \rho \in \mathbb{R}^+$, $\mathbf{P} + \rho\mathbf{T} \in \mathrm{Sym}_n^+$.*
- *if $\mathbf{T}$ has at least one negative eigenvalue, then for*

$$0 \leq \rho \leq -\frac{\min(\mathrm{sp}(\mathbf{P}))}{\min(\mathrm{sp}(\mathbf{T}))},$$

we have $\mathbf{P} + \rho\mathbf{T} \in \mathrm{Sym}_n^+$. $\mathrm{sp}(\cdot)$ stands for the spectrum (i.e. the set of (real) eigenvalues) of a (symmetric) matrix.

Proof.

- if $\mathbf{T}$ is positive semidefinite, $\forall \mathbf{x} \in \mathbb{R}^n$ we have

$$\mathbf{x}^T(\mathbf{P} + \rho\mathbf{T})\mathbf{x} = \mathbf{x}^T\mathbf{P}\mathbf{x} + \rho\mathbf{x}^T\mathbf{T}\mathbf{x} > 0,$$

since $\mathbf{x}^T\mathbf{T}\mathbf{x} \geq 0$. Then $\mathbf{P}+\rho\mathbf{T} \in \mathrm{Sym}_n^+$ for any $\rho \geq 0$.

- if $\mathbf{T}$ has at least one negative eigenvalue, we use a inequality due to Weyl (1912), [23, 24] bounding the eigenvalues of a sum of symmetric matrices by the sums of the extremal values of the eigenvalues of each matrix:

Lemma 6.2 (One of Weyl's inequalities [23])**.** *Let $\mathbf{N}, \mathbf{R} \in \mathrm{Sym}_n$, and $\mathbf{M} = \mathbf{R} + \mathbf{N}$. Then*

$$\min(\mathrm{sp}(\mathbf{M})) \leq \min(\mathrm{sp}(\mathbf{R})) + \min(\mathrm{sp}(\mathbf{N})).$$

Applying this inequality to $\mathbf{P} + \rho\mathbf{T}$, we get

$$\min(\mathrm{sp}(\mathbf{P} + \rho\mathbf{T})) \leq \min(\mathrm{sp}(\mathbf{P})) + \rho\min(\mathrm{sp}(\mathbf{T})).$$

If the right hand side is greater than 0, then $\mathbf{P} + \rho\mathbf{T} \in \mathrm{Sym}_n^+$. From this, isolating ρ, bearing in mind that $\min(\mathrm{Sp}(\mathbf{T})) < 0$, yields the result.

□

This theorem means that when the tangent vector at the given iterate happens to be positive semidefinite, then a classical Euler step will remain on the manifold. However, in the general case, the tangent vector may have negative eigenvalues, and then we can only guarantee that the next iterate will remain on the manifold for small enough time steps. Hence, large integration steps may lead the next iterate to leave the manifold.

6.2. Proof of the transitivity of the group action (11)

Proof. Let $\mathbf{X}, \mathbf{Y} \in \mathrm{Sym}_n^+$, and define $\mathbf{G} = \mathbf{Y}^{\frac{1}{2}}\mathbf{X}^{\frac{-1}{2}} \in GL_n(\mathbb{R})$. Then

$$\begin{aligned}\mathbf{\Lambda}(\mathbf{G}, \mathbf{X}) = \mathbf{G}\mathbf{X}\mathbf{G}^T &= \mathbf{Y}^{\frac{1}{2}}\mathbf{X}^{\frac{-1}{2}}\mathbf{X}(\mathbf{Y}^{\frac{1}{2}}\mathbf{X}^{\frac{-1}{2}})^T \\ &= \mathbf{Y}^{\frac{1}{2}}\mathbf{X}^{\frac{-1}{2}}\mathbf{X}^{\frac{1}{2}}\mathbf{X}^{\frac{1}{2}}\mathbf{X}^{\frac{-1}{2}}\mathbf{Y}^{\frac{1}{2}} \\ &= \mathbf{Y}^{\frac{1}{2}}\mathbf{I}\mathbf{Y}^{\frac{1}{2}} \\ &= \mathbf{Y}.\end{aligned}$$

□

6.3. Computation of the Lie algebra homomorphism (12)

We start from the definition given in Eq. (5), and use the fact that on a matrix group such as $GL_n(\mathbb{R})$, a curve $\boldsymbol{\rho}$ for which $\boldsymbol{\rho}(0) = \mathbf{I}$ has an initial speed that can be defined in the usual Euclidean sense, i.e. $\boldsymbol{\rho}'(0) = \mathbf{M}$, with $\mathbf{M} \in \mathcal{M}_n(\mathbb{R})$. Then, the curve can be written as a Taylor expansion at zero:

$$\boldsymbol{\rho}(s) = \mathbf{I} + s\mathbf{M} + o(s).$$

Then, working from the definition we have

$$\begin{aligned}\boldsymbol{\lambda}_*(\mathbf{M})(\mathbf{P}) &\triangleq \frac{d}{ds}\mathbf{\Lambda}(\boldsymbol{\rho}(s), \mathbf{P})|_{s=0} \\ &= \frac{d}{ds}(\boldsymbol{\rho}(s)\mathbf{P}\boldsymbol{\rho}(s)^T)|_{s=0} \\ &= \frac{d}{ds}((\mathbf{I} + s\mathbf{M} + o(s))\mathbf{P}(\mathbf{I} + s\mathbf{M} + o(s))^T)|_{s=0} \\ &= (\mathbf{M} + o(1))\mathbf{P}(\mathbf{I} + s\mathbf{M} + o(s)) \\ &\quad + (\mathbf{I} + s\mathbf{M} + o(s))^T\mathbf{P}(\mathbf{M} + o(1))^T|_{s=0} \\ &= \mathbf{M}\mathbf{P} + \mathbf{P}\mathbf{M}^T.\end{aligned}$$

Using similar arguments, we can easily check that the "product" (7) reduces to the usual matrix product in our case. We define $\mathbf{M} \in M_n(\mathbb{R})$ and $\mathbf{N} \in GL_n(\mathbb{R})$. We consider a curve $\boldsymbol{\rho} \in GL_n(\mathbb{R})$ as above (with initial speed $\mathbf{M}$). Then:

$$\begin{aligned}\frac{d}{ds}\boldsymbol{\rho}(s)\mathbf{N}|_{s=0} &= \frac{d}{ds}(\mathbf{I} + s\mathbf{M} + o(s))\mathbf{N} \\ &= (\mathbf{M}\mathbf{N} + o(1)\mathbf{N})|_{s=0} \\ &= \mathbf{M}\mathbf{N}.\end{aligned}$$

6.4. Derivation of the covariance equations and corresponding $\boldsymbol{\xi}$ for several equations of interest

Here we derive, for several examples of interest, equations followed by covariance matrices. The first three cases concern SDE on a stochastic process $\boldsymbol{X}_t \in \mathbb{R}^n$ for which we are interested in the evolution of the covariance matrix (we drop the time index for brevity)

$$\mathbf{P} = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T\right] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{X}]^T.$$

The last case is different in nature and concerns an optimal control problem in continuous time whose solution involves solving an equation on an SPD matrix.

6.4.1. Linear deterministic system with stochastic initial condition

The first equation of interest is simply a linear deterministic dynamical system, given by

$$\frac{d\boldsymbol{X}}{dt} = \mathbf{A}\boldsymbol{X}.$$

This describes a stochastic process if the initial condition is given by a probability distribution instead of a deterministic value. In that case, taking the expectation of the solutions for any possible $\boldsymbol{X}_0$, we get the same ODE on $\mathbf{m}(t) = \mathbb{E}[\boldsymbol{X}(t)]$.

$$\frac{d\mathbf{m}}{dt} = \mathbf{A}\mathbf{m}.$$

Then, to obtain a differential equation on $\mathbf{P}$, we first compute the derivative of $\boldsymbol{X}\boldsymbol{X}^T$:

$$\frac{d(\boldsymbol{X}\boldsymbol{X}^T)}{dt} = \frac{d\boldsymbol{X}}{dt}\boldsymbol{X}^T + \boldsymbol{X}\frac{d\boldsymbol{X}^T}{dt} = \mathbf{A}\boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{X}\boldsymbol{X}^T\mathbf{A}^T.$$

Then we can obtain, taking expectations in the previous equation, subtracting $\mathbf{m}\mathbf{m}^T$ to form the covariance (we can swap derivation and expectation by dominated convergence):

$$\begin{aligned}\frac{d\mathbf{P}}{dt} &= \frac{d}{dt}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right)\\ &= \mathbf{A}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] + \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\mathbf{A}^T - \mathbf{A}\mathbf{m}\mathbf{m}^T - \mathbf{m}\mathbf{m}^T\mathbf{A}^T\\ &= \mathbf{A}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right) + \left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right)\mathbf{A}^T\\ &= \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T.\end{aligned}$$

From this, it is clear that taking $\boldsymbol{\xi}(\mathbf{P}, t) \equiv \mathbf{A}$ in (13) corresponds exactly to this equation. In other words, the simplest ODE the group action (11) can handle models the evolution of the covariance matrix of a process passing through a linear dynamical system.

6.4.2. Multivariate Ornstein-Uhlenbeck process

We now switch to a an actual (Itô) SDE, in this case a linear one with a constant diffusion:

$$d\boldsymbol{X}_t = \mathbf{A}\boldsymbol{X}_t dt + \mathbf{B}d\boldsymbol{W}_t,$$

where $\boldsymbol{X} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times n}$, $\boldsymbol{W}_t \in \mathbb{R}^n$ is a multivariate Brownian motion (with independent entries). This is a multivariate generalization of the well-known Ornstein-Uhlenbeck process. Since the expectation of the Brownian motion is zero, the ODE followed by the mean of the process is the same as in the previous example. This entails that the mean of the process will converge to zero as well as long as $\mathbf{A}$ has no eigenvalues with a positive real part. There is a closed form solution for both the trajectories and the covariance matrix for a deterministic initial condition (see. e.g. [19]). Even if the initial condition is stochastic, we can derive the ODE governing the evolution of the covariance of the process. To do this, we must first derive the SDE followed by $\boldsymbol{X}\boldsymbol{X}^T$, using Itô's Lemma [18]. We have $(\boldsymbol{X}\boldsymbol{X}^T)_{ij} = X_iX_j$. Applying Itô's Lemma to this function yields:

$$\begin{aligned}d(X_iX_j) &= \sum_{k=1}^{n}\frac{\partial(X_iX_j)}{\partial X_k}dX_k + \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{\partial^2(X_iX_j)}{\partial X_k\partial X_l}dX_k dX_l\\ &= X_j dX_i + X_i dX_j + dX_i dX_j.\end{aligned}$$

Hence, gathering all these terms for $1 \le i, j \le n$ in a matrix form, we get:

$$d(\boldsymbol{X}\boldsymbol{X}^T) = (d\boldsymbol{X})\boldsymbol{X}^T + \boldsymbol{X}d\boldsymbol{X}^T + d\boldsymbol{X}d\boldsymbol{X}^T.$$

Replacing $d\boldsymbol{X}$ with its expression, and expanding, we obtain:

$$\begin{aligned}d(\boldsymbol{X}\boldsymbol{X}^T) &= (\mathbf{A}\boldsymbol{X}dt + \mathbf{B}d\boldsymbol{W}_t)\boldsymbol{X}^T + \boldsymbol{X}(\mathbf{A}\boldsymbol{X}dt + \mathbf{B}d\boldsymbol{W}_t)^T\\ &\qquad + (\mathbf{A}\boldsymbol{X}dt + \mathbf{B}d\boldsymbol{W}_t)(\mathbf{A}\boldsymbol{X}dt + \mathbf{B}d\boldsymbol{W}_t)^T\\ &= \mathbf{A}\boldsymbol{X}\boldsymbol{X}^T dt + \mathbf{B}d\boldsymbol{W}_t\boldsymbol{X}^T + \boldsymbol{X}\boldsymbol{X}^T\mathbf{A}^T dt\\ &+ \boldsymbol{X}d\boldsymbol{W}_t^T\mathbf{B}^T + \mathbf{A}\boldsymbol{X}\boldsymbol{X}^T\mathbf{A}^T dt^2 + \mathbf{A}\boldsymbol{X}dt d\boldsymbol{W}_t^T\mathbf{B}^T\\ &+ \mathbf{B}d\boldsymbol{W}_t\mathbf{A}\boldsymbol{X}dt + \mathbf{B}d\boldsymbol{W}_t d\boldsymbol{W}_t\mathbf{B}^T.\end{aligned}$$

Using the usual multiplication rules, $dt^2 = dtd\boldsymbol{W}_t = 0$ and $d\boldsymbol{W}_t d\boldsymbol{W}_t^T = dt\mathbf{I}$, we obtain

$$\begin{aligned}d(\boldsymbol{X}\boldsymbol{X}^T) &= (\mathbf{A}\boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{X}\boldsymbol{X}^T\mathbf{A}^T + \mathbf{B}\mathbf{B}^T)dt\\ &\qquad + \boldsymbol{X}d\boldsymbol{W}_t^T\mathbf{B}^T + \mathbf{B}d\boldsymbol{W}_t\boldsymbol{X}^T.\end{aligned}$$

Note that this is still an Itô SDE, since by vectorizing $\boldsymbol{X}\boldsymbol{X}^T$ into a n^2 dimensional vector, the last SDE can be rewrittten in the usual form. To do this, we use the well known property of the Kronecker product (denoted as $\otimes$): $\text{vec}(\mathbf{B}\mathbf{V}\mathbf{A}) = (\mathbf{A} \otimes \mathbf{B})\text{vec}(\mathbf{V})$, with vec the vectorization operator. This yields

$$\begin{aligned}d(\text{vec}(\boldsymbol{X}\boldsymbol{X}^T)) &= ((\boldsymbol{X} \otimes \mathbf{A} + \mathbf{A}^T \otimes \boldsymbol{X})\boldsymbol{X} + \text{vec}(\mathbf{B}\mathbf{B}^T))dt\\ &\qquad + (\mathbf{I} + \mathbf{K})(\boldsymbol{X} \otimes \mathbf{B})d\boldsymbol{W}_t,\end{aligned}$$

where $\mathbf{K} \in \mathbb{R}^{n^2\times n^2}$ is the commutator matrix, i.e. the matrix such that $\text{vec}(\mathbf{M}^T) = \mathbf{K}\text{vec}(\mathbf{M})$. Writing out the solution to this SDE, and taking expectations, we get:

$$\frac{d\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]}{dt} = \mathbf{A}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] + \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\mathbf{A}^T + \mathbf{B}\mathbf{B}^T.$$

Similarly as in the previous section, forming the covariance matrix yields:

$$\begin{aligned}\frac{d\mathbf{P}}{dt} &= \frac{d}{dt}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right)\\ &= \mathbf{A}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] + \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\mathbf{A}^T + \mathbf{B}\mathbf{B}^T\\ &\quad - \mathbf{A}\mathbf{m}\mathbf{m}^T - \mathbf{m}\mathbf{m}^T\mathbf{A}^T\\ &= \mathbf{A}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right) + \left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right)\mathbf{A}^T\\ &\quad + \mathbf{B}\mathbf{B}^T\\ &= \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T.\end{aligned}$$

Finally, by defining $\boldsymbol{\xi}(\mathbf{P}, t) = \mathbf{A} + \frac{1}{2}\mathbf{B}\mathbf{B}^T\mathbf{P}^{-1}$ in (13) we can use our framework to integrate this equation.

6.4.3. Multivariate Geometric Brownian Motion

The multivariate GBM SDE is given by Eq. (17). When $\mathbf{A}$ and $\mathbf{B}$ commute, and $\mathbf{A}+\frac{1}{2}\mathbf{B}^2$ has no eigenvalue with a negative real part, [17] provides a closed form solution for an initial value of $\mathbf{x}_0$ under the form:

$$\boldsymbol{X}(t) = \exp(t\mathbf{A} + \mathbf{B}W_t)\mathbf{x}_0.$$

With our choice of parameters (normal commuting matrices for $\mathbf{A}$ and $\mathbf{B}$, see Sec. 6.5), we can obtain [17] a closed form solution on the variance of the process at each time step, and for any initial (deterministic) $\mathbf{x}_0$: $\mathbb{E}[||\boldsymbol{X}(t)||^2]$:

$$\mathbb{E}[||\boldsymbol{X}(t)||^2] = \mathbb{E}[\boldsymbol{X}(t)^T\boldsymbol{X}(t)] = ||\exp(t\mathbf{Q})\mathbf{x}_0||^2,$$

where $\mathbf{Q} = \frac{\mathbf{A}+(\mathbf{B}+\mathbf{B}^T)^2}{4}$. Then, regardless of the initial distribution, the process converges to $\mathbf{0}$ (with zero covariance).

Letting $\boldsymbol{\theta} \triangleq \mathbf{A}+\frac{1}{2}\mathbf{B}^2$, the ODE followed by $\mathbf{m}(t)$ is

$$\frac{d\mathbf{m}}{dt} = \boldsymbol{\theta}\mathbf{m}.$$

To obtain the equation on the covariance matrix, we follow the same method as in the previous section, and from the expression of $d(\boldsymbol{X}\boldsymbol{X}^T)$ given by Itô's Lemma, we can compute:

$$\begin{aligned}
d(\boldsymbol{X}\boldsymbol{X}^T) &= (\boldsymbol{\theta}\boldsymbol{X}dt + \mathbf{B}\boldsymbol{X}dW_t)\boldsymbol{X}^T + \boldsymbol{X}(\boldsymbol{\theta}\boldsymbol{X}dt + \mathbf{B}\boldsymbol{X}dW_t)^T \\
&\quad + (\boldsymbol{\theta}\boldsymbol{X}dt + \mathbf{B}\boldsymbol{X}dW_t)(\boldsymbol{\theta}\boldsymbol{X}dt + \mathbf{B}\boldsymbol{X}dW_t)^T \\
&= \boldsymbol{\theta}\boldsymbol{X}\boldsymbol{X}^T dt + \mathbf{B}\boldsymbol{X}dW_t\boldsymbol{X}^T + \boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\theta}^T dt \\
&\quad + \boldsymbol{X}dW_t\boldsymbol{X}^T\mathbf{B}^T + \boldsymbol{\theta}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\theta}^T dt^2 + \boldsymbol{\theta}\boldsymbol{X}dtdW_t\boldsymbol{X}^T\mathbf{B}^T \\
&\quad + \mathbf{B}\boldsymbol{X}dW_t dt\boldsymbol{X}^T\boldsymbol{\theta}^T + \mathbf{B}\boldsymbol{X}dW_t^2\boldsymbol{X}^T\mathbf{B}^T,
\end{aligned}$$

which yields, after applying the conventions mentioned above:

$$\begin{aligned}
d(\boldsymbol{X}\boldsymbol{X}^T) &= (\boldsymbol{\theta}\boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\theta}^T + \mathbf{B}\boldsymbol{X}\boldsymbol{X}^T\mathbf{B}^T)dt \\
&\quad + (\mathbf{B}\boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{X}\boldsymbol{X}^T\mathbf{B}^T)dW_t.
\end{aligned}$$

Writing out the solution and taking the expectation:

$$\frac{d\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]}{dt} = \boldsymbol{\theta}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] + \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\boldsymbol{\theta}^T + \mathbf{B}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\mathbf{B}^T.$$

Then, forming the covariance matrix yields:

$$\begin{aligned}
\frac{d\mathbf{P}}{dt} &= \frac{d}{dt}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right) \\
&= \boldsymbol{\theta}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] + \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\boldsymbol{\theta}^T + \mathbf{B}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\mathbf{B}^T \\
&\quad - \boldsymbol{\theta}\mathbf{m}\mathbf{m}^T - \mathbf{m}\mathbf{m}^T\boldsymbol{\theta}^T \\
&= \boldsymbol{\theta}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right) + \left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbf{m}\mathbf{m}^T\right)\boldsymbol{\theta}^T \\
&\quad + \mathbf{B}\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\mathbf{B}^T \\
&= \boldsymbol{\theta}\mathbf{P} + \mathbf{P}\boldsymbol{\theta}^T + \mathbf{B}(\mathbf{P} + \mathbf{m}\mathbf{m}^T)\mathbf{B}^T,
\end{aligned}$$

which is indeed equal to Eq. 19. We notice that for the multivariate GBM, the ODEs on the mean and covariance are coupled, since the equation on the covariance involves the mean. In practice, we first integrate the equation on the mean and plug the solution into the covariance ODE.

Finally, choosing $\boldsymbol{\xi}$ as in Eq. (20) puts this equation within our integration framework.

6.4.4. Riccati differential equation in continuous time finite horizon linear quadratic optimal control

In this last application, we switch to a linear quadratic optimal control problem with finite horizon and continuous time [14]. We consider a state variable $\mathbf{x} \in \mathbb{R}^n$, initially at $\mathbf{x}(0) = \mathbf{x}_0$, that is subject to a linearly controlled linear dynamical system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u},$$

where $\mathbf{u} \in \mathbb{R}^n$ is the control variable, and $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$. We want to control the system, e.g. to stabilize the trajectories with $t \in [0, t_f]$, if $\mathbf{A}$ is such that without control, the system diverges. This will happen if $\mathbf{A}$ has at least one eigenvalue with a positive real part. In this case, the goal is then to minimize the quadratic cost function wrt to $\mathbf{u}$, leading to a so called Linear Quadratic Regulator (LQR):

$$\mathcal{L}(\mathbf{u}) = \frac{1}{2}\left(||\mathbf{x}(t_f)||^2_{\mathbf{Q}(t_f)} + \int_0^{t_f}(||\mathbf{x}||^2_{\mathbf{Q}} + ||\mathbf{u}||^2_{\mathbf{R}})dt\right),$$

depending on the choices of the SPD matrices $\mathbf{Q}_{t_f}$, $\mathbf{Q}$ and $\mathbf{R}$ to set a tradeoff between the energy of the control and how much we want to push the trajectory of $\mathbf{x}$ towards zero. We let $||\mathbf{x}||^2_{\boldsymbol{\Sigma}} = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^n$, and $\boldsymbol{\Sigma} \in \mathrm{Sym}_n^+$.

The closed-loop control solution is given in closed form by $\mathbf{u}(t) = \mathbf{K}(t)\mathbf{x}(t)$, where $\mathbf{K}(t)$ is a time dependent gain, given by

$$\mathbf{K}(t) = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}(t).$$

In the expression of the gain, $\mathbf{P}(t) \in \mathrm{Sym}_n^+$ is an SPD matrix that can be obtained by solving the so-called Riccati differential equation:

$$\frac{d\mathbf{P}}{dt} = -(\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{Q}),$$

with terminal condition $\mathbf{P}(t_f) = \mathbf{Q}_{t_f}$. This time varying algebraic Riccati equation can be derived from the cost function, either using Pontryagin's maximum principle or dynamic programming, through the Hamilton-Jacobi-Bellman equations [14].

Once again, our proposed framework can handle this differential equation (and its many variants) by choosing $\boldsymbol{\xi}(\mathbf{P}, t) = -\mathbf{A} + \frac{1}{2}\mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T - \frac{1}{2}\mathbf{Q}\mathbf{P}^{-1}$.

6.5. Choice of numerical values for the SDE (17)

For the case study of Section 4, we chose $n = 3$. We chose $\mathbf{B}$ to be a normal matrix (i.e. commuting with its transpose, to simplify computations, and so we can apply the spectral theorem):

$$\mathbf{B} = 0.25 \times \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

We chose $\mathbf{A}$ to commute with $\mathbf{B}$. For this, $\mathbf{A}$ and $\mathbf{B}$ must be simultaneously diagonalizable. Hence we chose:

$$\mathbf{A} = 0.25 \times \mathbf{U} \begin{bmatrix} -5+20i & 0 & 0 \\ 0 & -5-20i & 0 \\ 0 & 0 & -4 \end{bmatrix} \mathbf{U}^H,$$

where $\mathbf{U} \in \mathbb{C}^{n \times n}$ is a unitary (complex valued) matrix diagonalizing $\mathbf{B}$ (whose eigenvalues are, in order, $(1 \pm i\sqrt{3})/2$ and 2). We know such a unitary matrix exists thanks to the spectral theorem, that applies to normal matrices. ^{H} is the Hermitian transpose operator. $\mathbf{A}$ is actually a real valued matrix because its complex eigenvalues are conjugate. With these choices, we can check that $\mathbf{A} + \frac{1}{2}\mathbf{B}^2$ is negative definite, guaranteeing that $\boldsymbol{X}(t)$ converges to zero for $t \to \infty$.

6.6. Riemannian metrics on Sym_n^+

In this section, we simply give the expressions of the two Riemannian distances on Sym_n^+ used in this paper, i.e. the log Euclidean distance and the affine invariant distance. We refer to [9, 10, 20] for more details on the Riemannian metrics generating those distances.

The Log-Euclidean distance [9] is defined as

$$d_{LE}(\mathbf{P}_1, \mathbf{P}_2) = || \log(\mathbf{P}_1) - \log(\mathbf{P}_2) ||_F,$$

where $\log$ is the matrix (principal) logarithm and $|| \cdot ||_F$ is the Frobenius norm.

The affine-invariant (sometimes called Fisher-Rao) distance [10, 20] is defined as

$$d_A(\mathbf{P}_1, \mathbf{P}_2) = || \log(\mathbf{P}_1^{-1/2} \mathbf{P}_2 \mathbf{P}_1^{-1/2}) ||_F.$$

The corresponding Riemann exponential map $\mathrm{Exp}_{\mathbf{P}}^A$ at $\mathbf{P} \in \mathrm{Sym}_n^+$, used in the Riemannian-RK4 method can be obtained as as [10]:

$$\mathrm{Exp}_{\mathbf{P}}^A(\mathbf{\Sigma}) = \mathbf{P}^{1/2} \exp(\mathbf{P}^{-1/2} \mathbf{\Sigma} \mathbf{P}^{-1/2}) \mathbf{P}^{1/2}.$$

7. REFERENCES

[1] Dominic Jordan and Peter Smith, *Nonlinear ordinary differential equations: an introduction for scientists and engineers*, OUP Oxford, 2007.

[2] John Charles Butcher, *Numerical methods for ordinary differential equations*, John Wiley & Sons, 2016.

[3] Ernst Hairer, "Solving differential equations on manifolds," *Lecture Notes, Université de Geneve*, 2011.

[4] Arieh Iserles, Hans Z Munthe-Kaas, Syvert P Nørsett, and Antonella Zanna, "Lie-group methods," *Acta numerica*, vol. 9, pp. 215–365, 2000.

[5] P-A Absil, Robert Mahony, and Rodolphe Sepulchre, "Optimization algorithms on matrix manifolds," in *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[6] Uwe Helmke and John B. Moore, *Optimization and Dynamical Systems*, Springer London, 1996.

[7] Brynjulf Owren, "Lie group integrators," in *Discrete Mechanics, Geometric Integration and Lie–Butcher Series*, pp. 29–69. Springer, 2018.

[8] Elena Celledoni, Håkon Marthinsen, and Brynjulf Owren, "An introduction to lie group integrators–basics, new developments and applications," *Journal of Computational Physics*, vol. 257, pp. 1040–1061, 2014.

[9] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM journal on matrix analysis and applications*, vol. 29, no. 1, pp. 328–347, 2007.

[10] Rajendra Bhatia, "Positive definite matrices," in *Positive Definite Matrices*. Princeton university press, 2009.

[11] Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, and Yannick Berthoumieu, "Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1107–1116, 2017.

[12] Ian L Dryden, Alexey Koloydenko, and Diwei Zhou, "Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging," *The Annals of Applied Statistics*, vol. 3, no. 3, pp. 1102–1123, 2009.

[13] Petros Dellaportas and Mohsen Pourahmadi, "Large time-varying covariance matrices with applications to finance," *Technical Report, Athens University of Economics, Department of Statistics*, 2004.

[14] Tamer Başar and Geert Jan Olsder, *Dynamic noncooperative game theory*, SIAM, 1998.

[15] Geir Evensen, Femke C Vossepoel, and Peter Jan van Leeuwen, *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*, Springer Nature, 2022.

[16] Olivier Pannekoucke, Sophie Ricci, Sebastien Barthelemy, Richard Ménard, and Olivier Thual, "Parametric kalman filter for chemical transport models," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 68, no. 1, pp. 31547, 2016.

[17] G Barrera, MA Högele, and JC Pardo, "Cutoff stability of multivariate geometric brownian motion," *arXiv preprint arXiv:2207.01666*, 2022.

[18] Bernt Oksendal, *Stochastic differential equations: an introduction with applications*, Springer Science & Business Media, 2013.

[19] Attilio Meucci, "Review of statistical arbitrage, cointegration, and multivariate ornstein-uhlenbeck," *Cointegration, and Multivariate Ornstein-Uhlenbeck (May 14, 2009)*, 2009.

[20] Silvere Bonnabel and Rodolphe Sepulchre, "Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1055–1070, 2010.

[21] Noura Dridi, Lucas Drumetz, and Ronan Fablet, "Learning stochastic dynamical systems with neural networks mimicking the euler-maruyama scheme," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1990–1994.

[22] Duong Nguyen, Said Ouala, Lucas Drumetz, and Ronan Fablet, "Variational deep learning for the identification and reconstruction of chaotic and stochastic dynamical systems from noisy and partial observations," *arXiv preprint arXiv:2009.02296*, 2020.

[23] Hermann Weyl, "Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung)," *Mathematische Annalen*, vol. 71, no. 4, pp. 441–479, 1912.

[24] Terence Tao, "254A, Notes 3a: Eigenvalues and sums of Hermitian matrices," `https://terrytao.wordpress.com/2010/01/12/254a-notes-3a-eigenvalues-and-sums-of-hermitian-matrices/`, 2010, [Online; accessed 09-Oct-2022].