Solving SDEs on Stiefel manifolds with polar retractions

Du Nguyen

June 16, 2024

1 Solving SDEs on Stiefel manifolds with polar retractions

This note explains how to solve SDEs on Stiefel manifolds using polar retractions, using the methods of [NS24]. Alternative retractions, such as qr decomposition or Cayley transforms could also be used, although the analysis may be more complicated. We also show explicitly the SDE for certain Riemanian-Brownian motions satisfy the conditions in [MPS16].

For integers $n > p > 0$, the Stiefel manifold $\mathcal{M} = \mathrm{St}_{n,p}$ consists of orthogonal matrices $X \in \mathcal{E} := \mathbb{R}^{n\times p}$ satisfying the condition $X^{\mathsf{T}}X = \mathrm{I}_p$.

In the terminology of [NS24], the polar retraction is a $\mathcal{E}$-tubular retraction, mapping $Q \in \mathbb{R}^{n\times p}$ to $\pi(Q) = X := UV^{\mathsf{T}} \in \mathrm{St}_{n,p}$, where $Q = U\Sigma V^{\mathsf{T}}$ is a reduced SVD decomposition, Σ and V^{T} are in $\mathbb{R}^p$, with U and V are orthogonal matrices and Σ is diagonal. We can see $Q = XV\Sigma V^{\mathsf{T}}$, with X is orthogonal while $V\Sigma V^{\mathsf{T}}$ is symmetric. We also note

$$(V\Sigma V^{\mathsf{T}})^2 = Q^{\mathsf{T}}Q. \tag{1.1}$$

Thus, a result of the paper, section 7.2, is, if we are given an Itō (or Stratonovich) equation on $\mathbb{R}^{n\times p}$ having $\mathcal{M} = \mathrm{St}_{n,p}$ as an invariant manifold (as in proposition 1), we can just retract each $\mathcal{E}$-iteration step back to $\mathcal{M}$. The resulting simulation on $\mathcal{M}$ will have similar orders of accuracy.

Another method, proposed in the paper, using a modified drift is as follows. Let η be a tangent vector to $\mathcal{M}$ at X, thus, $(X^{\mathsf{T}}\eta)_{\mathrm{sym}} = 0$. The associated tangent-retraction is the map $\mathfrak{r}(X,\eta) = \pi(X+\eta)$, obtained by applying the polar decomposition to $X+\eta$ [AM12]. From eq. (1.1),

$$\pi(X+\eta) = (X+\eta)\{(X+\eta)^{\mathsf{T}}(X+\eta)\}^{-\frac{1}{2}} = (X+\eta)\{I+\eta^{\mathsf{T}}\eta\}^{-\frac{1}{2}},$$

which gives us the Taylor expansion

$$\mathfrak{r}(X,\eta) = X + \eta - \frac{1}{2}X\eta^{\mathsf{T}}\eta + O(|\eta|^3). \tag{1.2}$$

The method proposed in section 7.3 suggests retracting the stochastic vector

$$\mathring{\sigma}(X,t)\Delta_{W_t} + h\mu_{\mathfrak{r}}(X_t,t)$$

where Δ_{W_t} is drawn from a zero-drift normal distribution on $\mathcal{W} = \mathbb{R}^k$ with covariant $h\mathrm{I}_{\mathbb{R}^k}$, and the modified drift $\mu_{\mathfrak{r}} = \mu - \frac{1}{2}\sum_i \mathfrak{r}^{(2)}(x, \mathring{\sigma}w_i, \mathring{\sigma}w_i)$ is

$$\mu_{\mathfrak{r}} = \mu + X\{\frac{1}{2}\sum_i (\mathring{\sigma}w_i)^{\mathsf{T}}(\mathring{\sigma}w_i)\} \tag{1.3}$$

where $w_i, i = 1\cdots k$ is a basis of $\mathbb{R}^k$, using eq. (1.2)

As an example, for $\alpha > 0$ consider the following Riemannian Brownian SDE on $\mathrm{St}_{n,p}$ (equation 6.23 in [NS24])

$$dX_t = -(\frac{n-p}{2} + \frac{p-1}{4\alpha})X_t dt + \{(\mathrm{I}_n - X_t X_t^{\mathsf{T}})dW_t + \alpha^{-\frac{1}{2}}X_t(X_t^{\mathsf{T}}dW_t)_{\mathrm{skew}}\}. \tag{1.4}$$

Here, we take $\mathcal{W} = \mathcal{E}$, W_t is the Wiener process on $\mathcal{E}$. The equation describes the Riemannian Brownian motion on $\mathrm{St}_{n,p}$ corresponding to the metric described in [HMS21, Ngu23], for two tangent vectors ξ,η at $X \in \mathrm{St}_{n,p}$

$$\langle \xi,\eta\rangle_\alpha = \mathrm{Tr}(\xi^{\mathsf{T}}\eta) + (\alpha-1)\,\mathrm{Tr}(\xi^{\mathsf{T}}XX^{\mathsf{T}}\eta) \tag{1.5}$$

with $\alpha = 1$ corresponding to the embedded Riemannian metric, $\alpha = \frac{1}{2}$ corresponding to the *canonical metric*. For this equation,

$$\mathring{\sigma}(X)W = \mathring{\sigma}(X)_\alpha W := \{(\mathrm{I}_n - XX^{\mathsf{T}})W + \alpha^{-\frac{1}{2}}X(X^{\mathsf{T}}W)_{\mathrm{skew}}\}. \tag{1.6}$$

Here, $\dim\mathcal{W} = n \times p = k$, the basis $\{w_s\}_{s=1}^k$ could be taken to be the set of elementary matrices $\{E_{ij}\}_{i=1,j=1}^{i=n,j=p}$ (all entries of E_{ij} are zero except for entry (i,j) is 1).

In the notation of [MPS16], the stochastic component is defined by $k = n\times p$ vector fields $\mathbf{F}_s = \mathbf{F}_{ij} = \mathring{\sigma}(X)E_{ij}$ ([MPS16] uses d to denote the number of vector fields), and $\mathbf{F}_0$ in that citation is the drift

$$\mu_\alpha(X_t) = -(\frac{n-p}{2} + \frac{p-1}{4\alpha})X_t. \tag{1.7}$$

The requirement (8) of theorem 1 of that citation is clearly satisfied. The requirement (7) becomes

$$\sum_{ij}(\mathring{\sigma}_\alpha E_{ij})^{\mathsf{T}}(\mathring{\sigma}_\alpha E_{ij}) = (n-p+\frac{p-1}{2\alpha})\,\mathrm{I}_p, \tag{1.8}$$

which we will show instantly, as it also appears in eq. (1.3).

Using lemma 1 in [NS24] we have

$$\begin{aligned}\sum_{ij}(\mathring{\sigma}_\alpha E_{ij})^{\mathsf{T}}(\mathring{\sigma}_\alpha E_{ij}) &= \sum(E_{ij})^{\mathsf{T}}(\mathring{\sigma}_\alpha\mathring{\sigma}_\alpha^{\mathsf{T}} E_{ij})\\ &= \sum_{ij} E_{ji}\{(\mathrm{I}_n - XX^{\mathsf{T}})E_{ij} + \alpha^{-1}X(X^{\mathsf{T}}E_{ij})_{\mathrm{skew}}\}\\ &= \mathrm{Tr}(\mathrm{I}_n - XX^{\mathsf{T}})\,\mathrm{I}_p + \frac{1}{2\alpha}E_{ji}XX^{\mathsf{T}}E_{ij} - \frac{1}{2\alpha}E_{ji}XE_{ji}X\\ &= (n-p+\frac{p}{2\alpha})\,\mathrm{I}_p - \frac{1}{2\alpha}X^{\mathsf{T}}X = (n-p+\frac{p-1}{2\alpha})\,\mathrm{I}_p\,.\end{aligned}$$

Thus, the condition (7) of theorem 1 of [MPS16] is satisfied, and this shows the adjusted drift $\mu_{\mathfrak{r}}$ is also zero for all α in eq. (1.4). In general, if a SDE is driven by the same stochastic term $\mathring{\sigma}_\alpha$ for some $\alpha > 0$, then the adjusted drift is given as

$$\mu_{\mathfrak{r}} = \mu + \frac{1}{2}(n-p+\frac{p-1}{2\alpha})X_t = \mu - \mu_\alpha \tag{1.9}$$

Thus, for the equation $dX_t = \mu(X_t,t) + \sigma_\alpha(X_t)dW_t$, we propose two retractive versions of Eule-Maruyama using the polar decomposition π,

$$X_{i+1} = \pi(X_i + h\mu(X_i,t_i) + \sigma_\alpha(X_i)\Delta_h W) \tag{1.10}$$

$$X_{i+1} = \pi(X + h(\mu(X_i,t_i) - \mu_\alpha(X_i)) + \sigma_\alpha\Delta_h W) \tag{1.11}$$

while for the Riemannian-Brownian motion ($\mu = \mu_\alpha$), eq. (1.11) becomes $X_{i+1} = \pi(X_i + \xi_i)$ for $\xi_i = \sigma_\alpha(X_i)\Delta_h W$. In this case, the second-order retraction in the example after theorem 4 in [NS24] adds another iteration

$$X_{i+1} = \pi(X_i + \xi_i - (1-\alpha)(\xi_i - X_iX_i^{\mathsf{T}}\xi_i)\xi_i^{\mathsf{T}}X_i). \tag{1.12}$$

For a Stiefel manifold with this metric, we have a special situation where eq. (1.12) could be replaced by $X_{i+1} = \pi(X_i + \xi_i)$. This is because from equation (2) in [SHSW23], the extra term when using a retraction is proportional to $\int_{S_X}\frac{\nabla}{d\tau}\frac{d}{d\tau}Ret_x(\tau\xi)d\xi$, and the difference between extra terms in eq. (1.11) and eq. (1.12) is proportional to

$$\int_{S_x}(\xi - XX^{\mathsf{T}}\xi)\xi^{\mathsf{T}}Xd\xi,$$

which integrates to zero. To see this, at $X = \mathrm{I}_{n,p} = \begin{bmatrix} \mathrm{I}_p \\ 0_{(n-p)\times p} \end{bmatrix}$, set $\xi = \begin{bmatrix} A_\xi \\ B_\xi \end{bmatrix}$, with A_ξ is an antisymmetric matrix, the integral becomes

$$\int_{\xi \in S_{T_X \mathrm{St}_{n,p}}} B_\xi A_\xi d\xi.$$

A change of variable $J : A_\xi \mapsto -A_\xi$ will change the sign of $B_\xi A_\xi$, but the Jacobian adjustment $|\det Jac(J)|$ is one, thus the integral is zero.

We have also tested numerically that these iteration methods give similar results numerically.

References

[AM12] P. Absil and J. Malick. Projection-like retractions on matrix manifolds. *SIAM J. Optim*, 22(1):135–158, 2012.

[HMS21] K. Hüper, I. Markina, and F. Silva Leite. A Lagrangian approach to extremal curves on Stiefel manifolds. *Journal of Geometric Mechanics*, 13:55–72, 2021.

[MPS16] Goran Marjanovic, Marc J. Piggott, and Victor Solo. Numerical methods for stochastic differential equations in the Stiefel manifold made simple. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 2853–2860, 2016.

[Ngu23] D. Nguyen. Operator-valued formulas for Riemannian gradient and Hessian and families of tractable metrics in Riemannian optimization. *Journal of Optimization Theory and Applications*, 198:135–164, 2023.

[NS24] D. Nguyen and S Sommer. Second-order differential operators, stochastic differential equations and brownian motions on embedded manifolds. arXiv:2406.02879, 2024.

[SHSW23] S. Schwarz, M. Herrmann, A. Sturm, and M. Wardetzky. Efficient random walks on Riemannian manifolds. arXiv:2202.00959, 2023.