(will be inserted by the editor)

Newton retraction as approximate geodesics on submanifolds

Ruda Zhang

Abstract Efficient approximation of geodesics is crucial for practical algorithms on manifolds. Here we introduce a class of retractions on submanifolds, induced by a foliation of the ambient manifold. They match the projective retraction to the third order and thus match the exponential map to the second order. In particular, we show that Newton retraction (NR) is always stabler than the popular approach known as oblique projection or orthographic retraction (OR): per Kantorovich-type convergence theorems, the superlinear convergence regions of NR include those of OR. We also show that NR always has a lower computational cost; it can be twice as fast, and possibly better if constraints are sparse. The preferable properties of NR are useful for sampling, optimization, Bayesian inference, and many other statistical problems on manifolds.

Keywords Riemannian manifold · geodesic · exponential map · retraction · Newton's method

Mathematics Subject Classification (2010) 53-08 · 65D15

1 Introduction

Consider a function

$$F \in C^k(\mathbb{R}^n, \mathbb{R}^c),$$

for which 0 is a regular value. By the regular level set theorem (see e.g. Hirsch, 1976, Thm 3.2), the zero set $F^{-1}(0)$ is a properly embedded C^k submanifold of $\mathbb{R}^n$ with codimension c and dimension $d = n - c$. We call $F(x)$ the constraint function and define the constraint manifold

$$\mathcal{M} = F^{-1}(0).$$

Depending on the context, $\mathcal{M}$ may also be called an implicitly-defined manifold, a solution manifold, or an equilibrium manifold. A geodesic $\gamma_v(t)$ in the manifold is a curve with initial location x, initial velocity v, and zero intrinsic acceleration. All the geodesics are encoded in the exponential map $\exp : \mathscr{E} \mapsto \mathcal{M}$ such that

$$\exp(x, v) = \gamma_v(1),$$

where $(x, v) \in \mathscr{E} \subset T\mathcal{M}$.

The exponential map is crucial for analysis and computation on manifolds. Application to problems on manifolds include optimization (Adler et al., 2002; Absil et al., 2008; Ge et al., 2015; Zhang and Sra, 2016; Boumal et al., 2018), differential equations (Hairer et al., 2006), interpolation (Sander, 2015), sampling (Brubaker et al., 2012; Byrne and Girolami, 2013; Liu et al., 2016; Leimkuhler and Matthews, 2016; Zappa et al., 2018; Mangoubi and Smith, 2018; Lelievre et al., 2019; Goyal and Shetty, 2019; Zhang, 2020), approximate Bayesian computation (Marin et al., 2012; Chua, 2020), and many other problems in statistics (Chen, 2020). If the exponential map is not known in an analytic form or is not computationally tractable, it can be approximated by numerically integrating the geodesic trajectory, i.e. solving the ordinary differential equation $D_t\gamma_v'(t) = 0$ with initial conditions $\gamma_v(0) = x$, $\gamma_v'(0) = v$. For submanifolds, this can also be done by projecting $x + v$ to $\mathcal{M}$, which requires solving a constrained minimization problem. In general, an approximation to the exponential map is

This work was supported by the National Science Foundation grant DMS-1638521.

R. Zhang
The Statistical and Applied Mathematical Sciences Institute, Durham, NC
E-mail: rzhang@samsi.info
Also at Department of Mathematics, North Carolina State University, Raleigh, NC
E-mail: rzhang27@ncsu.edu

Table 1 Computation of geodesic steps on submanifolds, a qualitative comparison.

method	manifold	approximation	stepsize	cost	examples
analytic geodesics	simple	exact	any	-	Byrne and Girolami (2013); Liu et al. (2016); Mangoubi and Smith (2018); Goyal and Shetty (2019)
numerical geodesics	level set	variable order	tiny	high	Leimkuhler and Matthews (2016)
projective retraction	level set	2nd order	almost any	high	
orthographic retraction	level set	2nd order	small	low	Brubaker et al. (2012); Zappa et al. (2018); Lelievre et al. (2019)
retraction by foliation	general	2nd order (1)	large	-	Zhang (2020); Zhang and Ghanem (2020)
Newton retraction	level set	2nd order (1)	large (3)	lower (4)	Adler et al. (2002)

* Cross-referenced are results in this paper.

called a retraction (Adler et al., 2002). As is widely acknowledged (see e.g. Absil et al., 2008; Zhang and Sra, 2016; Boumal et al., 2018), retractions are often far less difficult to compute than the exponential map, which allows for practical and efficient algorithms.

In this article, we introduce a class of second-order retractions on submanifolds, which move points along manifolds orthogonal to the constraint manifold. We call them "retractions induced by normal foliation". Of particular interest among this class is Newton retraction (NR), which we show to have better convergence property and lower computational cost than the popular approach known as oblique projection or orthographic retraction (OR). Table 1 gives a qualitative overview of methods for computing geodesics, where we summarize some of our results.

1.1 Related Literature

Retraction was first defined in Adler et al. (2002) for Newton's method for root finding on submanifolds of Euclidean spaces, which was applied to a constrained optimization problem over a configuration space of the human spine. In particular, they proposed a retraction for constraint manifolds using Newton's method, which we study in this paper. Retractions are widely useful for optimization on manifolds. Noisy stochastic gradient method (Ge et al., 2015) escapes saddle points efficiently, which uses projective retraction for constrained problems. For geodesically convex optimization, Zhang and Sra (2016) studied several first-order methods using the exponential map and a projection oracle, while acknowledging the value of retractions. The Riemannian trust region method has a sharp global rate of convergence to an approximate second-order critical point, where any second-order retraction may be used (Boumal et al., 2018).

Sampling on manifolds often involves simulating a diffusion process, which is usually done by a Markov Chain Monte Carlo (MCMC) method. Brubaker et al. (2012) generalized Hamiltonian Monte Carlo (HMC) methods to distributions on constraint manifolds. Their constrained HMC simulates the Hamiltonian dynamics with the RATTLE scheme (Andersen, 1983), in which orthographic retraction is used to maintain state and momentum constraints. Byrne and Girolami (2013) proposed geodesic Monte Carlo (gMC), an HMC method for submanifolds with known geodesics. Liu et al. (2016) proposed two stochastic gradient MCMC methods for manifolds with known geodesics, including a variant of gMC. For molecular dynamic simulation with configuration constraints, Leimkuhler and Matthews (2016) proposed a numerical scheme for constrained Langevin dynamics, which samples the configuration manifold. Criticizing the stability and accuracy limitations in the SHAKE method and its RATTLE variant—both of which use orthographic retraction—their scheme approximated geodesics by numerical integration. Zappa et al. (2018) proposed reversible Metropolis random walks on constraint manifolds, which use orthographic retraction. Lelievre et al. (2019) generalized the previous work to constrained generalized HMC, allowing for gradient forces in proposal; it uses RATTLE. In this line of work, it is necessary to check that the proposal is actually reversible, because large timesteps can lead to bias in the invariant measure. The authors pointed out that numerical efficiency in these algorithms requires a balance between stepsize and the proportion of reversible proposed moves. More recently, Zhang (2020) proposed a family of ergodic diffusion processes for sampling on constraint manifolds, which use retractions defined by differential equations. To sample the uniform distribution on compact Riemannian manifolds, Mangoubi and Smith (2018) proposed geodesic walk, which uses the exponential map. Goyal and Shetty (2019) used a similar approach to sample the uniform distribution on compact, convex subsets of Riemannian manifolds, which can be adapted to sample an arbitrary density using a Metropolis filter; it uses the exponential map.

Table 2 Some MCMC methods on constraint manifolds.

paper	MCMC family	method name	geodesic computation
Brubaker et al. (2012)	HMC	constrained HMC	RATTLE (OR)
Byrne and Girolami (2013)	HMC	gMC	exponential map
Liu et al. (2016)	HMC	gSGNHT, SGGMC	exponential map
Leimkuhler and Matthews (2016)	Langevin	g-OBABO	numerical integration
Mangoubi and Smith (2018)	-	geodesic walk	exponential map
Goyal and Shetty (2019)	Metropolis	adapted geodesic walk	exponential map
Zappa et al. (2018)	Metropolis	Metropolis random walk	OR
Lelievre et al. (2019)	GHMC	constrained GHMC	RATTLE (OR)
Zhang (2020)	-	ergodic diffusions	retraction by ODE

Many other statistical problems on constraint manifolds are discussed in Chen (2020), where Newton retraction can be very useful. For example, in approximate Bayesian computation (Marin et al., 2012), constraints are imposed on certain summary statistics and one wants to approximate the posterior distribution. Newton retraction can be used to maintain such constraints on the parameter space.

In probabilistic learning on manifolds (Soize et al., 2020; Zhang et al., 2020), a retraction based on constrained gradient flow is used for inference and data augmentation on density ridge (Zhang and Ghanem, 2020), which falls into the family of retractions defined in this paper.

2 Newton Retraction

2.1 Preliminaries

Retractions (Adler et al., 2002) are mappings that preserve the correct initial location and velocity of the geodesics; they approximate the exponential map to the first order. Second-order retractions (Absil et al., 2008, Prop 5.33) are retractions with zero initial acceleration; they approximate the exponential map to the second order. In general, we define retraction of an arbitrary order as follows.

Definition 1 *Retraction* $R(x,v)$ of order i on a C^k manifold, $1 \leq i < k$, is a C^{k-1} mapping to the manifold from an open subset of the tangent bundle containing all the zero tangent vectors, such that at every zero tangent vector it agrees with the exponential map in Riemannian distance to the i-th order:

$$R \in C^{k-1}(U, M), \zeta(M) \subset U \subset T\mathcal{M},$$
$$\forall (x,v) \in T\mathcal{M}, t \in \mathbb{R},$$
$$R(x,tv) = \exp(x,tv) + o(t^i),$$

in the sense that,

$$\lim_{t\to 0} d_g(R(x,tv), \exp(x,tv))/t^i = 0$$

Absil and Malick (2012) defined a class of retractions on submanifolds of Euclidean spaces, called projection-like retractions. This includes projective and orthographic retractions, both of which are second-order.

Definition 2 (Absil and Malick (2012), Def 14) *Retractor* $V(x,v)$ of a C^k submanifold of $\mathbb{R}^n$ with codimension c is a C^{k-1} mapping from tangent vectors to linear c-subspaces of the ambient space, such that for every zero tangent vector, affine space of the form

$$A(x,v) = x + v + V(x,v)$$

intersects the submanifold transversely:

$$V \in C^{k-1}(U, G_{c,n}),$$
$$\forall (x,v) \in U, A(x,v) \cap \mathcal{M} \neq \emptyset,$$
$$\forall x \in \mathcal{M}, A(x,0) \pitchfork \mathcal{M}.$$

Here $G_{c,n}$ is the Grassmann manifold.
Projection-like retraction $R_V(x,v)$ induced by a retractor is a correspondence that takes a tangent vector to the set of points closest to the origin of the affine space that intersects the submanifold:

$$R_V : U \rightrightarrows \mathcal{M},$$
$$R_V(x,v) = \underset{y \in A(x,v) \cap \mathcal{M}}{\arg\min} \|y - (x+v)\|.$$

In particular, it is a mapping if the tangent vector is small enough:

$$\forall x \in \mathcal{M}, \exists U' \subset U, (x,0) \in U' :$$
$$R_V|_{U'} \in C^{k-1}(U', \mathcal{M}).$$

Definition 3 *Projective retraction*

$$R_P(x,v) = P_{\mathcal{M}}(x+v),$$

where projection

$$P_{\mathcal{M}}(x) = \arg\min\{\|y - x\| : y \in \mathcal{M}\},$$

and can be seen as the projection-like retraction induced by retractor

$$V(x,v) = N_p\mathcal{M},$$
$$p = P_{\mathcal{M}}(x+v).$$

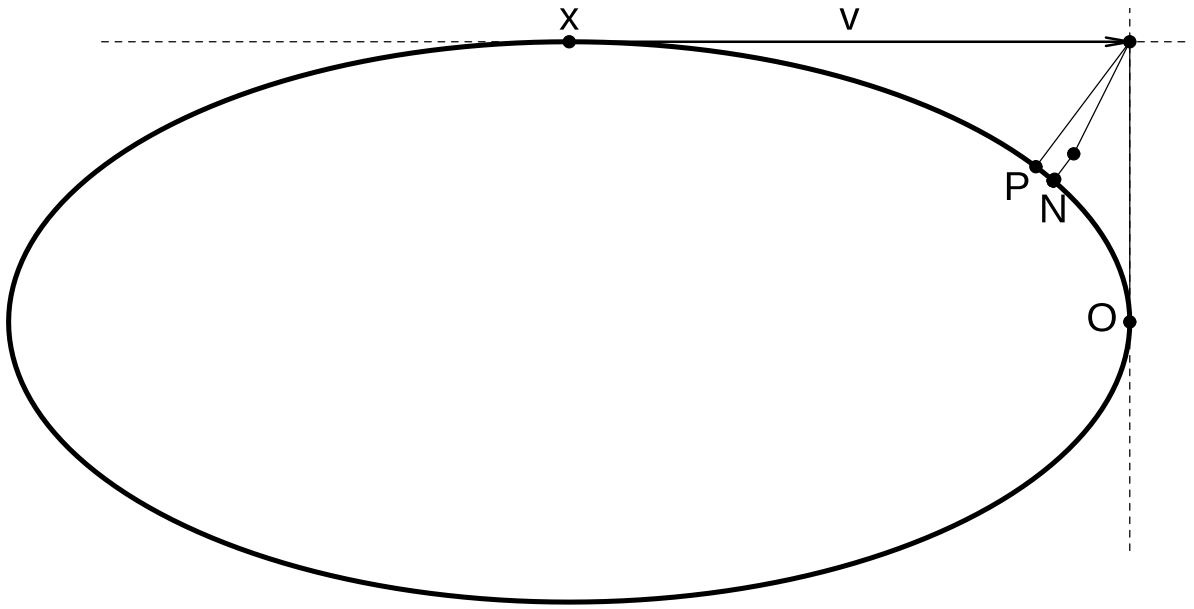

Fig. 1 Illustration of retractions on an ellipse. With initial point x and tangent vector v, retraction $R(x, v)$ returns: O, orthographic; P, projective; N, Newton (intermediate points are shown).

Orthographic retraction $R_O(x, v)$ is the projection-like retraction induced by retractor

$$V(x, v) = N_x\mathcal{M}.$$

Lemma 1 (Absil and Malick (2012), Thm 15, 22) *Projection-like retraction is a (first-order) retraction. It is second-order if the retractor maps every zero tangent to the normal space:*

$$\forall x \in \mathcal{M}, V(x, 0) = N_x\mathcal{M}.$$

2.2 Main Results

Here we define another class of second-order retractions on submanifolds, based on foliations that intersect the submanifold orthogonally. Such foliations may be generated by a dynamical system, where each leaf is a stable invariant manifold. In particular, we are interested in Newton retraction, generated by a discrete-time dynamical system with quadratic local convergence. Such retractions can also be generated by flows: Zhang (2020) used gradient flows of squared 2-norm of the constraint, Zhang and Ghanem (2020) used constrained gradient flows of log density function; both have linear local convergence and, with sufficiently small steps, global convergence.

Definition 4 *Normal foliation* $\mathscr{F}$ of a neighborhood of a codimension-c submanifold of a Riemannian n-manifold is a partition of the neighborhood into connected c-submanifolds (called the *leaves* of the foliation) which intersect the submanifold orthogonally:

$$\mathscr{F} = \sqcup_{p\in\mathcal{M}}\mathscr{F}_p,$$
$$\cup_{p\in\mathcal{M}}\mathscr{F}_p = D, \mathcal{M} \subset D \subset \widetilde{\mathcal{M}},$$
$$\forall p \in \mathcal{M}, T_p\mathscr{F}_p = (T_p\mathcal{M})^\perp$$

Retraction induced by a normal foliation is the map

$$R_{\mathscr{F}} = \pi \circ \widetilde{R},$$

where $\widetilde{R}$ is a retraction on the ambient manifold and π is the canonical projection:

$$\pi : D \mapsto \mathcal{M},$$
$$\forall x \in D, x \in \mathscr{F}_{\pi(x)}.$$

If $\widetilde{\mathcal{M}} = \mathbb{R}^n$, let $\widetilde{R}$ be the Euclidean exponential map

$$E(p, v) = p + v,$$

we have

$$R_{\mathscr{F}}(x, v) = \pi(x + v),$$
$$R_{\mathscr{F}} : E^{-1}(D) \mapsto \mathcal{M}.$$

Recall that for the under-determined system of nonlinear equations $F(x) = 0$, Newton's minimum-norm step is

$$\delta(x) = -J^\dagger(x)F(x),$$

where Jacobian $J = \nabla F$, $J_{ij} = \partial F^i/\partial x^j$, and $\dagger$ denotes the Moore-Penrose inverse. If J has full row rank, then

$$J^\dagger = J^T(JJ^T)^{-1}.$$

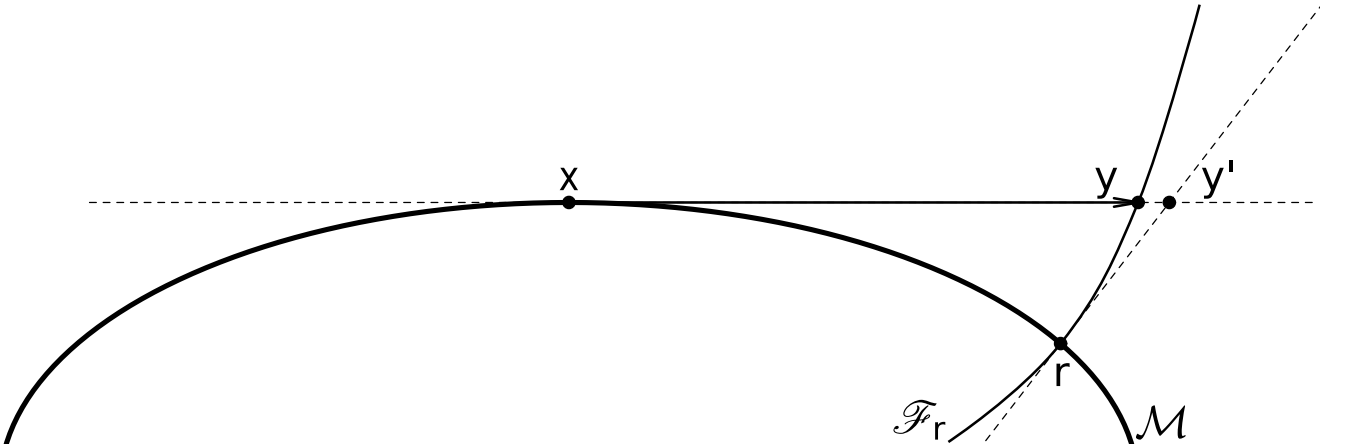

Fig. 2 Approximate projection by a normal foliation. Leaf $\mathscr{F}_r$ intersects the submanifold $\mathcal{M}$ orthogonally at r and intersects tangent space $T_x\mathcal{M}$ at y. Tangent spaces $T_r\mathscr{F}_r$ and $T_x\mathcal{M}$ intersect at y'.

Newton map, or Newton's least-change update, is

$$N_F(x) = x + \delta(x) = x - J^\dagger(x)F(x).$$

Newton limit map is the mapping

$$N_F^\infty \in C^{k-1}(D, \mathcal{M}),$$
$$N_F^\infty(x) = \lim_{n\to\infty} N_F^n(x),$$

where D is a neighborhood of $\mathcal{M}$. Adler et al. (2002, Ex 4) showed that given any retraction $\widetilde{R}$ on the ambient manifold, $N_F^\infty \circ \widetilde{R}$ is a retraction on $\mathcal{M}$. We call this map Newton retraction.

Definition 5 *Newton retraction* is the map

$$R_N \in C^{k-1}(E^{-1}(D), \mathcal{M}),$$
$$R_N(x, v) = N_F^\infty(x + v),$$

and can be seen as the retraction induced by the *normal foliation determined by the Newton map*:

$$\mathscr{N} = \sqcup_{p\in\mathcal{M}} \mathscr{N}_p,$$
$$\mathscr{N}_p = \{x \in D : N_F^\infty(x) = p\}.$$

To show that the retractions we defined are second-order, we give two lemmas. First, projective retractions form a class of retractions of order two and not of any higher order. Second, the retraction induced by a normal foliation matches the projective retraction at zero tangent vectors to the third order.

Lemma 2 *For every C^k submanifold $\mathcal{M}$, $k \geq 3$, the projective retraction R_P satisfies:*

$$\forall (x, v) \in T\mathcal{M}, t \in \mathbb{R},$$
$$R_P(x, tv) = \exp(x, tv) + o(t^2).$$

There exists a C^k submanifold, $k \geq 4$, such that the previous condition does not hold if $o(t^2)$ is replaced by $o(t^3)$.

Lemma 3 *Given a submanifold $\mathcal{M} \subset \mathbb{R}^n$ and a normal foliation $\mathscr{F}$ of a neighborhood of $\mathcal{M}$,*

$$\forall (x, v) \in T\mathcal{M},$$
$$R_{\mathscr{F}}(x, v) = R_P(x, v) + o(\|v\|^3).$$

Theorem 1 *Retraction $R_{\mathscr{F}}$ induced by a normal foliation, which includes Newton retraction R_N, is a second-order retraction. This characterization of order is sharp.*

Although R_N and R_O are both second-order retractions, they have different domain sizes. Notice that the projection from a Euclidean space onto a compact subset is uniquely defined except for a subset of measure zero, so the projective retraction R_P on a compact submanifold is defined almost everywhere on the tangent bundle. On the other hand, R_O may be undefined for large tangent vectors as the affine space $x + v + N_x\mathcal{M}$ fails to intersect the submanifold, see Figure 1. This precludes the domain of R_O to a relatively small subset of the tangent bundle, regardless of implementation. Since R_N matches R_P to the third order while R_O and R_P can differ on the third order, it is easy to see that R_N can have a larger domain than R_O.

Now we characterize the domain of R_N relative to that of R_O in their usual implementation. Algorithm 1 gives an implementation of R_N, which uses Newton's method to solve:

$$F(x) = 0$$
$$x_0 = x + v \in \mathbb{R}^n.$$

In comparison, R_O is usually implemented using Newton's method to solve:

$$F(x_0 + J(x)^T y) = 0$$
$$y_0 = 0 \in \mathbb{R}^c$$

Algorithm 1 Newton Retraction

1: **Given**: point and tangent vector (x, v), convergence threshold c_0
2: $x \leftarrow x + v$
3: **repeat**
4: $\quad J \leftarrow J(x)$
5: $\quad$ solve $(JJ^T)y = F(x)$
6: $\quad \delta \leftarrow -J^T y$
7: $\quad x \leftarrow x + \delta$
8: **until** $\|\delta\| < c_0$
9: **return** x

This means replacing line 5 with

$$(JJ_{-1}^T)y = F(x)$$

and line 6 with

$$\delta \leftarrow -J_{-1}^T y,$$

where $J_{-1} = J(x)$ is evaluated at the input x. It can be seen as an augmented Jacobian algorithm for solving under-determined systems of nonlinear equations: denote the Stiefel manifold

$$V_{d,n} = \{X \in M_{n,d}(\mathbb{R}) : X^T X = I_d\},$$

given $V \in V_{d,n}$, step $\delta(x)$ is defined by

$$\begin{aligned} J(x)\delta(x) &= -F(x), \\ V\delta(x) &= 0. \end{aligned}$$

For R_O, the algorithm starts with $x_0 = x+v$ and V satisfies $J(x)V = 0$. Kantorovich-type convergence theorems for Newton's method and augmented Jacobian algorithms are given in Walker and Watson (1990), which provide sufficient conditions for immediately superlinear convergence.

Definition 6 Let $F \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ and $J = \nabla F$. Let $C \subset \mathbb{R}^n$ be open and convex, $\alpha \in (0, 1]$, $K \geq 0$, and $B > 0$. We say function F satisfies the *normal flow hypothesis*,

$$F \in \mathcal{H}_{\mathrm{NF}}(C; \alpha, K, B),$$

if $\forall x, y \in C$,

$$\begin{aligned} \|J(x) - J(y)\| &\leq K\|x - y\|^\alpha \quad (1) \\ \operatorname{rank}(J(x)) &= m \quad (2) \\ \|J(x)^\dagger\| &\leq B \quad (3) \end{aligned}$$

Given $V \in V_{d,n}$, we say function F satisfies the *augmented Jacobian hypothesis*,

$$F \in \mathcal{H}_{\mathrm{AJ}}(V, C; \alpha, K, B)$$

if $\forall x, y \in C$,

$$\begin{aligned} \|J(x) - J(y)\| &\leq K\|x - y\|^\alpha \quad (1) \\ \operatorname{rank}\begin{bmatrix} J(x) \\ V \end{bmatrix} &= n \quad (2') \\ \left\| \begin{bmatrix} J(x) \\ V \end{bmatrix}^{-1} \right\| &\leq B \quad (3') \end{aligned}$$

Theorem 2 (Walker and Watson (1990), Thm 2.1, 3.2)
If $F \in \mathcal{H}_{NF}(C; \alpha, K, B)$ then $\forall \eta > 0$, $\exists \epsilon > 0$, Newton's method satisfies:

$$\begin{aligned} \forall x_0 \in \{x \in C : B_\eta(x) \subset C, \|F(x)\| < \epsilon\}, \\ \exists \zeta \in C \cap F^{-1}(0), \lim_{k\to\infty} x_k = \zeta; \end{aligned}$$

in particular, $\exists \beta > 0$, $\forall k \in \mathbb{N}$,

$$\|x_{k+1} - \zeta\| \leq \beta \|x_k - \zeta\|^{1+\alpha}.$$

If $F \in \mathcal{H}_{AJ}(V, C; \alpha, K, B)$, then the previous statement holds for the augmented Jacobian algorithm.

Per the previous Kantorovich-type convergence theorem, it follows immediately from the next lemma that Newton retraction is always stabler than orthographic retraction. Recall that R_N has domain $E^{-1}(D)$, where D is the domain of N_F^∞, i.e. the convergence region of Newton's method, and R_O has domain U. Let $U' \subset U$ be the convergence region of the usual implementation of R_O.

Lemma 4 *For any $V \in V_{d,n}$, if*

$$F \in \mathcal{H}_{AJ}(V, C; \alpha, K, B),$$

then

$$F \in \mathcal{H}_{NF}(C; \alpha, K, B).$$

Theorem 3 *With the usual implementation of R_N and R_O, for any $\alpha \in (0, 1]$, the order-$(1 + \alpha)$ convergence region of R_O guaranteed by Theorem 2 is a subset of*

that of R_N:
Let

$$D_\alpha = \{x \in D : \exists(C, K, B, \eta, \epsilon), F \in \mathcal{H}_{NF}(C; \alpha, K, B), B_\eta(x) \subset C, \|F(x)\| < \epsilon\}$$

and

$$U'_\alpha = \{(x, v) \in U' : \exists(C, K, B, \eta, \epsilon), F \in \mathcal{H}_{AJ}(V_x, C; \alpha, K, B), B_\eta(x + v) \subset C, \|F(x + v)\| < \epsilon\},$$

where $V_x \in V_{d,n}$, $J(x)V_x = 0$, then

$$\forall \alpha \in (0, 1], U'_\alpha \subset E^{-1}(D_\alpha).$$

Our next result shows that Newton retraction is always faster than orthographic retraction.

Theorem 4 *With the usual implementation of R_N and R_O, for any $\alpha \in (0, 1]$, for any $(x, v) \in U'_\alpha$, the number of operations required for R_N to converge is no greater than that for R_O. In particular, R_N admits linear solvers about twice as efficient as those for R_O.*

3 Discussion

We focus on comparing two methods for computing geodesics, Newton retraction (NR) and orthographic retraction (OR), despite other methods listed in Table 1. This is because for general constraint manifolds the exponential map is not available, and in cases where it is available, retractions are often much easier to compute. For algorithms that compute geodesics by numerical integration (Leimkuhler and Matthews, 2016), although they can have higher orders of approximation depending on the scheme, it is at the cost of smaller stepsizes. Projective retraction R_P comes up in theoretical papers (Ciccotti et al., 2008, Eq. 3.1) but is rarely used in practice due to the optimization problem involved. Retractions $R_{\mathscr{F}}$ induced by normal foliation typically have linear convergence rather than quadratic convergence, especially those implemented by solving ordinary differential equations (ODEs) (Zhang, 2020; Zhang and Ghanem, 2020). This leave us with R_N and R_O.

We have shown that, with small stepsizes, R_N has the same order of approximation to the exponential map as R_O (Lemma 1 and Theorem 1). And with their usual implementation, R_N has a larger domain than R_O (Theorem 3) which means larger stepsizes can be used, and R_N is always faster to compute than R_O regardless of stepsize (Theorem 4). Moreover, R_N can be seen as an efficient approximation of R_P (Lemma 3).

For applications that take small geodesic steps, such as molecular dynamics, NR can be readily applied instead of OR due to faster speed. (See Table 2 for some uses in MCMC.) The computational complexities per iteration in NR and OR (see Algorithm 1) are both dominated by the evaluation of the Jacobian which are $c \times n$ real-valued functions, and the linear solver in use, which invokes $\mathcal{O}(c^3)$ algebraic operations. Both methods terminate after a fixed number of iterations, because of their immediately superlinear and typically quadratic convergence. But because the coefficient matrix in NR is symmetric positive-definite, the linear equations can be solved using Cholesky decomposition, which is roughly twice as fast as LU decomposition. Cholesky decomposition is also numerically more stable and saves about half the storage, which is significant when c is large. If the coefficient matrix is large and sparse, efficient iterative methods such as preconditioned conjugate gradient can be used for NR, which can save more computation time and storage.

In case Jacobian evaluation is expensive and high numerical accuracy is unnecessary, one may consider a modified Newton retraction (mNR): run Algorithm 1 line 4 only for the first iteration, denote the outcome as J_0, and replace line 5 with

$$(J_0 J_0^T)y = F(x).$$

As a corollary of Lemma 1, mNR is a second-order retraction. In this context, a natural implementation of R_O is to use a chord method: remove line 4, and replace line 5 with

$$(J_{-1} J_{-1}^T)y = F(x).$$

Both methods have linear convergence, but mNR has a faster rate: let

$$\|x_{k+1} - x_\infty\| \leq \mu \|x_k - x_\infty\|,$$

then exists $\lambda \in (0, 1)$ such that $\mu = \lambda \|v\|^q$, where $q = 1$ for R_O and $q = 2$ for mNR (see e.g. Allgower and Georg, 1990, Eq 6.2.15). By Lemma 4, however, mNR is no more stable than NR.

For applications where large stepsizes are desirable, NR has the extra advantage that it can still converge when OR cannot. On the other hand, OR can benefit from its linear geometric structure for specific uses. Zappa et al. (2018) used Metropolis adjustment to guarantee reversibility of random walks by OR, which eliminates bias in the invariant measure, without computing second-order derivatives. Lelievre et al. (2019) generalized this idea to GHMC. If R_P is used instead, reverse steps can still be exactly computed, but because the volume ratios depend on curvature and do not cancel out, second-order derivatives need to be computed. If NR is

used, exact reversal is not available due to nonlinearity of the normal foliation, but sampling bias can still be partially corrected by computing volume ratios in the forward step and an approximate reversal. We note that, with the timesteps for which the reversible methods are most efficient, most rejections are due to OR failing to converge rather than reverse projection check or Metropolis adjustment (see Lelievre et al., 2019, Table 1). Moreover, if the computation is approximate in nature, such as approximating a Bayesian posterior, exact sampling is unnecessary.

4 Numerical Experiments

In this section we quantify the properties of NR versus OR with some specific examples.

4.1 Compute time

First we compare their compute time per iteration, for a family of submanifolds of varying dimensions. Following Zappa et al. (2018), we use the orthogonal groups $\mathrm{O}(m)$, each of which consists of $m \times m$ orthogonal matrices:

$$\mathrm{O}(m) = \{x \in \mathbb{R}^{m\times m} : x^T x - I_m = 0\}.$$

To simplify discussion, we do not use the special orthogonal groups $\mathrm{SO}(m)$, which is one of the two connected components of $\mathrm{O}(m)$ where the matrices have determinant 1. Considering $\mathrm{O}(m)$ as a submanifold of $\mathbb{R}^{m\times m}$, it can be written as the constraint manifold

$$\begin{aligned}\mathrm{O}(m) &= F^{-1}(0)\\ F(x) &= (F_{ij}(x))_{i\le j}\\ F_{ij}(x) &= x_{ki}x_{kj} - \delta_{ij}\end{aligned}$$

In this case, we have ambient dimension $n = m^2$, codimension $c = m(m+1)/2$, and manifold dimension $d = m(m-1)/2$. The Jacobian of the constraint functions can be written as

$$J_{ijkl}(x) = \frac{\partial F_{ij}(x)}{\partial x_{kl}} = \delta_{lj}x_{ki} + \delta_{il}x_{kj}.$$

We note that the Jacobian is a sparse matrix: among the $c \times n$ entries, only m^3 entries can be non-zero, that is $2/(m+1)$ full.

To estimate average compute time, we obtain random points on $O(m)$ by generating random matrices with standard Gaussian entries:

$$y \in \mathbb{R}^{m\times m}, y_{ij} \sim N(0,1),$$

and let x be the Q of the QR decomposition of y. We obtain a random tangent vector at each x by generating random anti-symmetric matrices with standard Gaussian entries:

$$\begin{aligned}\Omega \in \mathbb{R}^{m\times m}, \Omega = -\Omega^T\\ \Omega_{ij} \sim N(0,1), i < j,\end{aligned}$$

and let $v = \sigma x\Omega$ where $\sigma = 0.01\sqrt{m/d}$.

We measure compute time of the linear algebra part (line 5) of the first iteration in NR and OR, which is the dominant part in this problem. For $m \le 30$, measurements are averaged over 10^3 runs; for $m > 30$, 100 runs are used. Figure 3a shows the relative time per iteration of NR. Consistent with Theorem 4, NR is faster than OR, and can be about twice as efficient using dense matrix algorithms. We also used sparse matrix algorithms for both methods, which are faster at higher dimensions. In this case, NR can be more than ten times as efficient.

4.2 Convergence region

Next we compare their convergence regions for two low-dimensional submanifolds. We note that NR typically can converge for very large step sizes. However, the outcome would be meaningless other than being on the constraint manifold. A better proxy for convergence region is the region of approximate zeros (see e.g. Blum et al., 1998, Sec 8.1). If x_0 is an approximate zero for an iterative root-finding method, the algorithm should produce a sequence $\{x_k\}_{k\in\mathbb{N}}$ the converges to a zero ζ, $F(\zeta) = 0$, such that relative distances to the associated zero is bounded by the quadratically convergent sequence with coefficient $1/2$:

$$\begin{aligned}a_k = \frac{\|x_k - \zeta\|}{\|x_0 - \zeta\|} \le b_k, k \in \mathbb{N}\\ b_0 = 1, b_{k+1} = b_k^2/2.\end{aligned}$$

An ellipse can be defined by the constraint function

$$F(x,y) = \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 - 1.$$

We take $a = 1$, $b = 0.5$. Using the parametrization

$$(x,y) = (a\cos\theta, b\sin\theta),$$

the tangent vectors can be written as

$$\left\{\frac{vt}{\|t\|} : t = \left(\frac{\sin\theta}{b}, \frac{-\cos\theta}{a}\right), \theta \in [0, 2\pi), v \in \mathbb{R}\right\}$$

Figure 3b shows the region of approximate zeros on the (θ, v)-plane. Consistent with Theorem 3, the convergence region of NR contains that of OR, and for each θ the relative size ranges between 2.17 and 2.64.

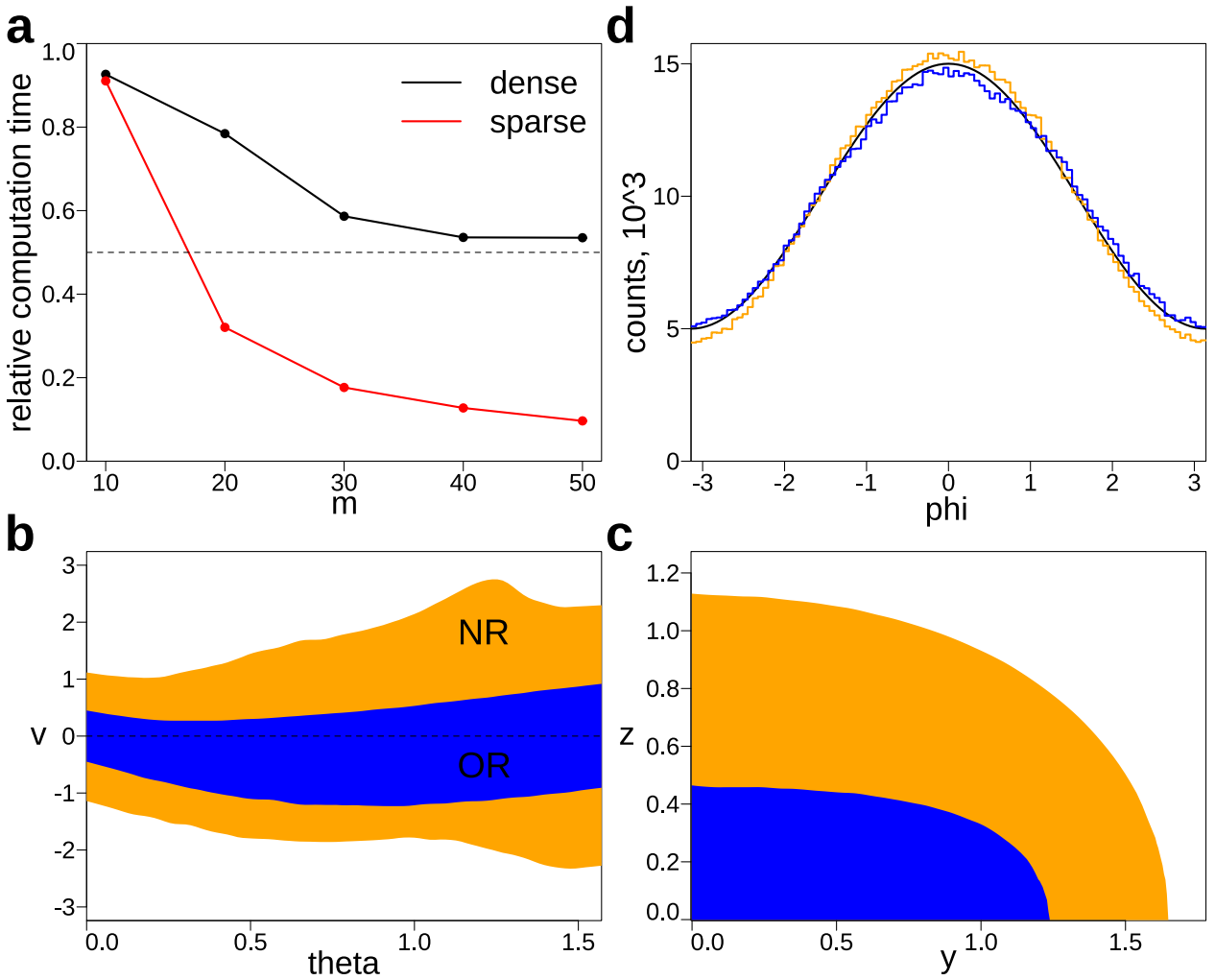

Fig. 3 Numerical results. (**a**) Relative compute time per iteration of NR, compared with OR, for orthogonal groups $\mathrm{O}(m)$. Both methods are computed with dense and sparse matrix algorithms for solving linear equations. (**b**) Location θ and tangent vector v on an ellipse that correspond to approximate zeros using NR and OR. (**c**) Same as (**b**) but for the 2-torus example, starting from $(R + r, 0, 0)$ with tangent vector $(0, y, z)$. (**d**) Histograms of angle ϕ of the samples using NR and OR. $N = 10^6$, $s = 0.1$. Theoretical density is shown in black line.

A 2-torus can be defined by the constraint function

$$F(x, y, z) = \left(\sqrt{x^2 + y^2} - R\right)^2 + z^2 - r^2.$$

We take $R = 1$, $r = 0.5$. Figure 3c shows the region of approximate zeros starting from $(1.5, 0, 0)$ with tangent vector $(0, y, z)$. The results are similar, and the relative size of the region for NR is about 3.1.

We note that, although the regions in these examples are simple, in general they are fractal, disconnected, and not simply connected.

4.3 MCMC sampling

Following Diaconis et al. (2013); Zappa et al. (2018); Lelievre et al. (2019), we sample the uniform distribution on the 2-torus, where the theoretical distribution of the parameters are know: ϕ has probability density function

$$f(\phi) = \frac{1}{2\pi}\left(1 + \frac{r}{R}\cos\phi\right).$$

We use the following random walk: given a point x_n on the torus, generate a random tangent vector v_n, which is an isotropic Gaussian vector with mean 0 and standard deviation s; use a retraction to find a point x_{n+1} on the torus; if the algorithm fails to converge, let $x_{n+1} = x_n$. Without Metropolis adjustment, finite timestep s leads to bias in the invariant measure (see Discussion). Figure 3d shows that, for the torus example, when the timestep results in small bias using OR, the bias is comparable if using NR.

5 Proofs

Proof of Lemma 2. Absil and Malick (2012, Ex 18) showed that projective retractions are second-order retractions, so we only need to show that the projective retraction of some manifold is exactly second-order. Consider the circle $\mathbb{S}^1$ as a submanifold of the Euclidean plane, identified with the set

$$\{(\cos\theta, \sin\theta) : \theta \in [0, 2\pi)\}.$$

Its exponential map is

$$\exp(x, v) = xe^{i\|v\|},$$

and its projective retraction is

$$R_P(x, v) = (x + v)/\|x + v\|.$$

Without loss of generality, consider the point $x = (1, 0)$ and tangent vectors $v = (0, \theta)$. Now we can write the exponential map as

$$\exp(x, v) = e^{i\theta}$$

and the projective retraction as

$$R_P(x, v) = e^{i \arctan\theta}.$$

So the distance between them is

$$d(\exp, R_P) = 2\sin((\theta - \arctan\theta)/2),$$

which has a Taylor expansion at zero as

$$d(\exp, R_P) = \theta^3/3 + O(\theta^5).$$

We can see that, for the unit circle, the projective retraction matches the exponential map up to the second order at zero tangent vectors. □

Proof of Lemma 3. For all $(x, v) \in T\mathcal{M}$ such that $y = x + v$ is in the neighborhood of $\mathcal{M}$ that is partitioned by $\mathscr{F}$, define $r \in \mathcal{M}$ to be the unique point such that $y \in \mathscr{F}_r$. Note that

$$T_r\mathscr{F}_r = N_r\mathcal{M},$$

so if $v = 0$ then $T_r\mathscr{F}_r$ and $T_x\mathcal{M}$ are orthogonal complements. Assume v is small enough such that affine spaces $r + T_r\mathscr{F}_r$ and $x + T_x\mathcal{M}$ intersect transversely, define the unique point

$$y' = (r + T_r\mathscr{F}_r) \cap (x + T_x\mathcal{M}).$$

These constructs are illustrated in Figure 2. Furthermore, define

$$v' = y' - x \in T_x\mathcal{M}$$
$$u' = y' - r \in N_r\mathcal{M}.$$

Because

$$R_{\mathscr{F}}(x, v) = R_P(x, v'),$$

then there exists $w \in T_x\mathcal{M}$ such that

$$R_{\mathscr{F}}(x, v) = R_P(x, v) + \left(\frac{\partial R_P}{\partial v}(x, w)\right)(v' - v),$$

that is,

$$R_{\mathscr{F}}(x, v) = R_P(x, v) + O(\|v' - v\|).$$

To prove the theorem, we only need to show

$$\|v' - v\| = O(\|v\|^4).$$

First we show that

$$\|u'\| = O(\|v'\|^2).$$

Parameterize $\mathcal{M}$ at x as the graph of a function

$$G : S_x \mapsto N_x\mathcal{M}, S_x \subset T_x\mathcal{M},$$
$$\forall v \in S_x, x + v + G(v) \in \mathcal{M}.$$

We see that

$$\|G(v')\| = O(\|v'\|^2).$$

Let

$$p' = x + v' + G(v'),$$

because

$$d(y', \mathcal{M}) \le d(y', p'),$$

we have

$$\|u'\| \le \|G(v')\|.$$

Thus,

$$\|u'\| = O(\|v'\|^2).$$

Second, we show that

$$\|v' - v\| = O(\|u'\|^2).$$

Parameterize $\mathscr{F}_r$ at r as the graph of a function

$$L : S_r \mapsto N_r\mathscr{F}_r, S_r \subset T_r\mathscr{F}_r,$$
$$\forall u \in S_r, r + u + L(u) \in \mathscr{F}_r.$$

We see that

$$\|L(u)\| = O(\|u\|^2).$$

For all $v, w \in \mathbb{R}^n$, if $\|v\|\|w\| \neq 0$, define angle

$$\angle(v, w) = \begin{cases} \arccos\left(\langle v, w\rangle/(\|v\|\|w\|)\right), & \|v\|\|w\| \neq 0 \\ \pi/2, & \text{otherwise.} \end{cases}$$

The first principal angle between two linear subspaces is defined as: let $V, W \subset \mathbb{R}^n$, then

$$\theta_1(V, W) = \min\{\angle(v, w) : v \in V, w \in W\}.$$

Because

$$\theta_1(T_x\mathcal{M}, N_x\mathcal{M}) = \pi/2,$$

we have

$$\theta_1(T_x\mathcal{M}, N_r\mathcal{M}) = \pi/2 + O(\|v\|).$$

Let

$$\beta = \angle(v - v', u - u'),$$

since

$$v - v' \in T_x\mathcal{M},$$
$$u - u' \in T_r\mathscr{F}_r = N_r\mathcal{M},$$

we have

$$\begin{aligned}\beta &\leq \theta_1(T_x\mathcal{M}, N_r\mathcal{M}) \\ &= \pi/2 + O(\|v\|).\end{aligned}$$

Thus,

$$\begin{aligned}\|v - v'\| &= (\sin\beta)^{-1}\|L(u)\| \\ &= O(\|L(u)\|) \\ &= O(\|u\|^2)\end{aligned}$$

and

$$\begin{aligned}\|u - u'\| &= \|v - v'\|\cos\beta \\ &= \|v - v'\|O(\|v\|).\end{aligned}$$

Because

$$\|u\| \leq \|u'\| + \|u' - u\|,$$

we have

$$\begin{aligned}\|u\| &\leq \|u'\| + o(\|v - v'\|) \\ &= \|u'\| + o(\|u\|^2),\end{aligned}$$

that is,

$$\|u\| = O(\|u'\|).$$

We conclude that

$$\|v' - v\| = O(\|u'\|^2).$$

Combining the previous two results gives

$$\|v' - v\| = O(\|v'\|^4).$$

Because

$$\|v'\| \leq \|v' - v\| + \|v\|,$$

we have

$$\|v'\| = O(\|v\|).$$

This means

$$\|v' - v\| = O(\|v\|^4).$$

□

Proof of Theorem 1. Combining Lemmas 2 and 3, we have

$$\begin{aligned}R_{\mathscr{F}}(x, tv) &= \exp(x, tv) + o(t^2) + o(t^3) \\ &= \exp(x, tv) + o(t^2),\end{aligned}$$

i.e. $R_{\mathscr{F}}$ is a second-order retraction. Beyn (1993, Thm 3.1) showed that N_F^∞ induces a foliation of a neighborhood of $\mathcal{M}$ into C^k c-submanifolds, which intersect $\mathcal{M}$ orthogonally. So R_N fits Definition 4 as a retraction induced by a normal foliation, and thus it is second-order. Recall the circle example in the proof of Lemma 2, if $\mathbb{S}^1$ is defined as the zero set of

$$F(x) = \|x\|^2 - 1, x \in \mathbb{R}^2,$$

then $R_N = R_P$, which means in this case R_N is only a second-order retraction. □

Proof of Lemma 4. By the definitions of the hypotheses, part (1) are identical, part (2) follows immediately from part (2'), so we only need to show part (3). For the rest of the proof, x is an arbitrary point in C. To simplify notation, we will ignore explicit dependence on x, such that J refers to $J(x)$, and so on. Since J has full rank, we have QR decomposition

$$J^T = QR,$$

where $Q \in V_{c,n}$ and $R \in U_+(c)$ is a upper triangular matrix of order c with positive diagonal entries. Let

$$\bar{Q} = [Q, \tilde{Q}] \in \mathrm{O}(n)$$

be an orthogonal matrix of order n, whose first c columns matches Q. Since $V \in V_{d,n}$, let

$$\begin{aligned}\tilde{Q}_0 &= V^T \\ \bar{Q}_0 &= [Q_0, \tilde{Q}_0] \in \mathrm{O}(n).\end{aligned}$$

First we show that Q^TQ_0 is non-singular and its spectral norm is no greater than 1. By (2'),

$$\begin{bmatrix} J \\ V \end{bmatrix} = \begin{bmatrix} R^TQ^T \\ \tilde{Q}_0^T \end{bmatrix}$$

is non-singular. Since R is invertible, this means $\begin{bmatrix} Q^T \\ \tilde{Q}_0^T \end{bmatrix}$ is non-singular, and thus

$$\begin{bmatrix} Q^T \\ \tilde{Q}_0^T \end{bmatrix} \bar{Q}_0 = \begin{bmatrix} Q^T \\ \tilde{Q}_0^T \end{bmatrix} [Q_0, \tilde{Q}_0]$$

is non-singular. Note that $\tilde{Q}_0^T Q_0 = 0$, $\tilde{Q}_0^T \tilde{Q}_0 = I_d$, so $\begin{bmatrix} Q^T Q_0 & Q^T \tilde{Q}_0 \\ 0 & I_d \end{bmatrix}$ is non-singular, which means $Q^T Q_0$ is non-singular. Moreover, let $u \in \mathbb{S}^c$, then

$$\begin{aligned} \|Q^T Q_0 u\| &\leq \left\| \begin{bmatrix} Q^T \\ \tilde{Q}^T \end{bmatrix} Q_0 u \right\| \\ &= \|\bar{Q}^T Q_0 u\| \\ &= \|Q_0 u\| \\ &= 1, \end{aligned}$$

which means

$$\rho(Q^T Q_0) \leq 1.$$

Now we prove (3). By (3'),

$$\left\| \begin{bmatrix} J \\ V \end{bmatrix}^{-1} \right\| \leq B,$$

that is, $\forall v \in \mathbb{R}^n$,

$$\left\| \begin{bmatrix} J \\ V \end{bmatrix}^{-1} v \right\| \leq B\|v\|.$$

This means, $\forall w \in \mathbb{R}^n$,

$$\|w\| \leq B \left\| \begin{bmatrix} J \\ V \end{bmatrix} w \right\|.$$

Equivalently, $\forall w \in \mathbb{S}^n$,

$$\left\| \begin{bmatrix} J \\ V \end{bmatrix} w \right\| \geq \frac{1}{B}.$$

Because $\forall u \in \mathbb{S}^c$, $Q_0 u \in \mathbb{S}^n$, so the previous inequality holds for $w = Q_0 u$. Note that

$$V Q_0 u = \tilde{Q}_0^T Q_0 u = 0,$$

the inequality becomes

$$\|J Q_0 u\| = \|R^T Q^T Q_0 u\| \geq \frac{1}{B}.$$

We have shown that $Q^T Q_0$ is non-singular and non-expansive, so

$$\forall z \in \mathbb{S}^c, \|R^T z\| \geq \frac{1}{B},$$

or equivalently,

$$\forall u \in \mathbb{R}^c, \|R^T u\| \geq \frac{1}{B}\|u\|.$$

Define

$$\tilde{u} = R^{-1} u,$$

then $\forall \tilde{u} \in \mathbb{R}^c$,

$$\|R^T R \tilde{u}\| \geq \frac{1}{B}\|R\tilde{u}\|.$$

Define

$$\bar{u} = R^T R \tilde{u},$$

then $\forall \bar{u} \in \mathbb{R}^c$,

$$\|R(R^T R)^{-1}\bar{u}\| \leq B\|\bar{u}\|.$$

The left-hand side equals $\|QR(R^T Q^T Q R)^{-1}\bar{u}\|$, that is $\|J^T (J J^T)^{-1}\bar{u}\|$ and in short $\|J^\dagger \bar{u}\|$. We conclude that

$$\|J^\dagger\| \leq B.$$

□

Proof of Theorem 3. For all $(x, v) \in U'_\alpha$, there exists $(C, K, B, \eta, \epsilon)$, such that

$$F \in \mathcal{H}_{\mathrm{AJ}}(V_x, C; \alpha, K, B).$$

By Lemma 4, we have

$$F \in \mathcal{H}_{\mathrm{NF}}(C; \alpha, K, B).$$

Therefore, $x + v \in D_\alpha$ and $(x, v) \in E^{-1}(D_\alpha)$. □

Proof of Theorem 4. Since $U'_\alpha \subset E^{-1}(D_\alpha)$, the update sequences $\{x_k\}_{k\in\mathbb{N}}$ in R_N and R_O both satisfy

$$\|x_{k+1} - \zeta\| \leq \beta \|x_k - \zeta\|^{1+\alpha}.$$

Because

$$d(x_0, \mathcal{M}) \leq d(x_0, \mathcal{M} \cap (x_0 + N_x \mathcal{M})),$$

the x_k in R_N will be remain closer to $\mathcal{M}$ than that in R_O after the same number of iterations, and thus R_N converges in no more iterations than R_O.

Moreover, at each iteration, R_N and R_O both evaluate $F(x)$ and $J(x)$, and solve an order-c system of linear equations, see line 5. But the coefficient matrix for R_O is a generic matrix $J J_{-1}^T$, while that for R_N is a symmetric positive-definite matrix $J J^T$, which admits triangular storage and faster linear solvers. The matrix product $J J_{-1}^T$ takes about $2nc^2$ floating point operations (FLOPs), half for multiplications and half for additions, while the cross product $J J^T$ takes about nc^2 FLOPs. LU decomposition of an order-c matrix takes about $2c^3/3$ FLOPs, while Cholesky decomposition for symmetric positive-definite matrices takes about $c^3/3$ FLOPs (Trefethen and Bau, 1997, pp. 175-177). Besides, Cholesky decomposition is always stable, which saves the need for pivoting operations. Therefore, about half of the algebraic computation can be saved in each iteration.

In conclusion, the overall number of operations in R_N is no greater than that in R_O. □

Acknowledgements The author thanks Mansoor Haider and Ryan Murray of NCSU, Greg Forest, Michael E Taylor, and Jeremy Louis Marzuola of UNC Chapel Hill, and Ernest Fokoue of Rochester Institute of Technology for support and valuable comments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Absil PA, Malick J (2012) Projection-like retractions on matrix manifolds. SIAM Journal on Optimization 22(1):135–158, DOI 10.1137/100802529

Absil PA, Mahony R, Sepulchre R (2008) Optimization Algorithms on Matrix Manifolds. Princeton University Press

Adler RL, Dedieu JP, Margulies JY, Martens M, Shub M (2002) Newton's method on riemannian manifolds and a geometric model for the human spine. IMA Journal of Numerical Analysis 22(3):359–390, DOI 10.1093/imanum/22.3.359

Allgower EL, Georg K (1990) Numerical Continuation Methods: An Introduction. Springer, DOI 10.1007/978-3-642-61257-2

Andersen HC (1983) Rattle: A velocity version of the shake algorithm for molecular dynamics calculations. Journal of Computational Physics 52(1):24–34, DOI 10.1016/0021-9991(83)90014-1

Beyn WJ (1993) On smoothness and invariance properties of the gauss-newton method. Numerical Functional Analysis and Optimization 14(5-6):503–514, DOI 10.1080/01630569308816536

Blum L, Cucker F, Shub M, Smale S (1998) Complexity and Real Computation. Springer, New York

Boumal N, Absil PA, Cartis C (2018) Global rates of convergence for nonconvex optimization on manifolds. IMA Journal of Numerical Analysis 39(1):1–33, DOI 10.1093/imanum/drx080

Brubaker M, Salzmann M, Urtasun R (2012) A family of MCMC methods on implicitly defined manifolds. In: Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, vol 22, pp 161–172, URL http://proceedings.mlr.press/v22/brubaker12.html

Byrne S, Girolami M (2013) Geodesic monte carlo on embedded manifolds. Scandinavian Journal of Statistics 40(4):825–845, DOI 10.1111/sjos.12036

Chen YC (2020) Solution manifold and its statistical applications. arXiv, URL https://arxiv.org/abs/2002.05297

Chua AJK (2020) Sampling from manifold-restricted distributions using tangent bundle projections. Statistics and Computing 30(3):587–602, DOI 10.1007/s11222-019-09907-8

Ciccotti G, Lelievre T, Vanden-Eijnden E (2008) Projection of diffusions on submanifolds: Application to mean force computation. Communications on Pure and Applied Mathematics 61(3):371–408, DOI 10.1002/cpa.20210

Diaconis P, Holmes S, Shahshahani M (2013) Sampling from a manifold. In: Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton, vol 10, Institute of Mathematical Statistics, Beachwood, Ohio, USA, pp 102–125, DOI 10.1214/12-IMSCOLL1006

Ge R, Huang F, Jin C, Yuan Y (2015) Escaping from saddle points — online stochastic gradient for tensor decomposition. In: Proceedings of The 28th Conference on Learning Theory, vol 40, pp 797–842, URL http://proceedings.mlr.press/v40/Ge15

Goyal N, Shetty A (2019) Sampling and optimization on convex sets in riemannian manifolds of non-negative curvature. In: Proceedings of the Thirty-Second Conference on Learning Theory, pp 1519–1561, URL http://proceedings.mlr.press/v99/goyal19a.html

Hairer E, Wanner G, Lubich C (2006) Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations. Springer, DOI 10.1007/3-540-30666-8

Hirsch MW (1976) Differential Topology. Springer, DOI 10.1007/978-1-4684-9449-5

Leimkuhler B, Matthews C (2016) Efficient molecular dynamics using geodesic integration and solvent-solute splitting. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 472(2189):20160138, DOI 10.1098/rspa.2016.0138

Lelievre T, Rousset M, Stoltz G (2019) Hybrid monte carlo methods for sampling probability measures on submanifolds. Numerische Mathematik 143(2):379–421, DOI 10.1007/s00211-019-01056-4

Liu C, Zhu J, Song Y (2016) Stochastic gradient geodesic mcmc methods. In: Advances in Neural Information Processing Systems 29, pp 3009–3017, URL http://papers.nips.cc/paper/6282-stochastic-gradient-geodesic-mcmc-methods

Mangoubi O, Smith A (2018) Rapid mixing of geodesic walks on manifolds with positive curvature. Annals of Applied Probability 28(4):2501–2543, DOI 10.1214/17-AAP1365

Marin JM, Pudlo P, Robert CP, Ryder RJ (2012) Approximate bayesian computational methods. Statistics and Computing 22(6):1167–1180, DOI 10.1007/s11222-011-9288-2

Sander O (2015) Geodesic finite elements of higher order. IMA Journal of Numerical Analysis 36(1):238–266, DOI 10.1093/imanum/drv016

Soize C, Ghanem RG, Desceliers C (2020) Sampling of bayesian posteriors with a non-gaussian probabilistic

learning on manifolds from a small dataset. Statistics and Computing DOI 10.1007/s11222-020-09954-6

Trefethen LN, Bau D (1997) Numerical Linear Algebra. SIAM

Walker HF, Watson LT (1990) Least-change secant update methods for underdetermined systems. SIAM Journal on Numerical Analysis 27(5):1227–1262, DOI 10.1137/0727071

Zappa E, Holmes-Cerfon M, Goodman J (2018) Monte carlo on manifolds: Sampling densities and integrating functions. Communications on Pure and Applied Mathematics 71(12):2609–2647, DOI 10.1002/cpa.21783

Zhang H, Sra S (2016) First-order methods for geodesically convex optimization. In: 29th Annual Conference on Learning Theory, vol 49, pp 1617–1638, URL `http://proceedings.mlr.press/v49/zhang16b.html`

Zhang R, Ghanem R (2020) Normal-bundle bootstrap. manuscript

Zhang R, Wingo P, Duran R, Rose K, Bauer J, Ghanem R (2020) Environmental economics and uncertainty: Review and a machine learning outlook. Oxford Encyclopedia of Environmental Economics DOI 10.1093/acrefore/9780199389414.013.572

Zhang W (2020) Ergodic sdes on submanifolds and related numerical sampling schemes. ESAIM: Mathematical Modelling and Numerical Analysis 54(2):391–430, DOI 10.1051/m2an/2019071