DISSIPATIVE NUMERICAL SCHEMES ON RIEMANNIAN MANIFOLDS WITH APPLICATIONS TO GRADIENT FLOWS[*]

ELENA CELLEDONI[†], SØLVE EIDNES[†], BRYNJULF OWREN[†],
AND TORBJØRN RINGHOLM[†]

Abstract. This paper concerns an extension of discrete gradient methods to finite-dimensional Riemannian manifolds termed discrete Riemannian gradients, and their application to dissipative ordinary differential equations. This includes Riemannian gradient flow systems which occur naturally in optimization problems. The Itoh–Abe discrete gradient is formulated and applied to gradient systems, yielding a derivative-free optimization algorithm. The algorithm is tested on two eigenvalue problems and two problems from manifold valued imaging: interferometric synthetic aperture radar denoising and diffusion tensor imaging denoising.

1. Introduction. When designing and applying numerical schemes for solving systems of ODEs and PDEs there are several important properties which serve to distinguish schemes, one of which is the preservation of geometric features of the original system. The field of geometric integration encompasses many types of numerical schemes for ODEs and PDEs specifically designed to preserve one or more such geometric features; a nonexhaustive list of features includes symmetry, symplecticity, first integrals (or energy), orthogonality, and manifold structures such as Lie group structure [14]. Energy conserving methods have a successful history in the field of numerical integration of ODEs and PDEs. In a similar vein, numerical schemes with guaranteed dissipation are useful for solving dissipative equations such as gradient systems.

As seen in [17], any Runge–Kutta method can be dissipative when applied to gradient systems as long as step sizes are chosen small enough; less severe but still restrictive conditions for dissipation in Runge–Kutta methods are presented in [13]. In [10], Gonzalez introduces the notion of discrete gradient schemes with energy preserving properties, later expanded upon to include dissipative systems in [21]. These articles consider ODEs in Euclidian spaces only with the exception of [13], where the authors also consider Runge–Kutta methods on manifolds defined by constraints. Unlike the Runge–Kutta methods, discrete gradient methods are dissipative for all step sizes, meaning one can employ adaptive time steps while retaining convergence toward fixed points [25]. However, one may experience a practical step size restriction when applying discrete gradient methods to very stiff problems, due to the lack of L-stability as seen when applying the Gonzalez and mean value discrete gradients to

[†]Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway (elena.celledoni@ntnu.no, solve.eidnes@ntnu.no, brynjulf.owren@ntnu.no, torbjorn.ringholm@ntnu.no).

problems with quadratic potentials [13], [15]. Motivated by their work on Lie group methods, the energy conserving discrete gradient method was generalized to ODEs on manifolds, and Lie groups particularly, in [7], where the authors introduce the concept of discrete differentials. In [5], this concept is specialized in the setting of Riemannian manifolds. To the best of our knowledge, the discrete gradient methods have not yet been formulated for dissipative ODEs on manifolds. Doing so is the central purpose of this article.

One of the main reasons for generalizing discrete gradient methods to dissipative systems on manifolds is that gradient systems are dissipative, and gradient flows are natural tools for optimization problems which arise in, e.g., manifold-valued image processing and eigenvalue problems. The goal is then to find one or more stationary points of the gradient flow of a functional $V : M \to \mathbb{R}$, which correspond to critical points of V. This approach is, among other optimization methods, presented in [1]. Since gradient systems occur naturally on Riemannian manifolds, it is natural to develop our schemes in a Riemannian manifold setting.

A similarity between the optimization algorithms in [1] and the manifold valued discrete gradient methods in [7] is their use of retraction mappings. Retraction mappings were introduced for numerical methods in [26] (see also [2]); they are intended as computationally efficient alternatives to parallel transport on manifolds. Our methods will be formulated as a framework using general discrete gradients on general Riemannian manifolds with general retractions. We will consider a number of specific examples that illustrate how to apply the procedure in practical problems.

As detailed in [11] and [22], using the Itoh–Abe discrete gradient [18], one can obtain an optimization scheme for n-dimensional problems with a limited degree of implicitness. At every iteration, one needs to solve n decoupled scalar nonlinear subequations, amounting to $\mathcal{O}(n)$ operations per step. In other discrete gradient schemes a system of n coupled nonlinear equations must be solved per iteration, amounting to $\mathcal{O}(n^2)$ operations per step. The Itoh–Abe discrete gradient method therefore appears to be well suited to large-scale problems such as image analysis problems, and so it seems natural to apply our new methods to image analysis problems on manifolds; see subsection 4.2. In [7], the authors generalize the average vector field [16] and midpoint [10] discrete gradients, but not the Itoh–Abe discrete gradient, to Lie groups and homogeneous manifolds. A novelty of this article is the formulation of the Itoh–Abe discrete gradient for problems on manifolds.

As examples we will consider two eigenvalue finding problems, in addition to the more involved problems of denoising interferometric synthetic aperture radar (InSAR) and diffusion tensor imaging (DTI) images using total variation (TV) regularization [30]. The latter two problems we consider as real applications of the algorithm. The two eigenvalue problems are included mostly for the exposition and illustration of our methods, as well as for testing convergence properties.

The paper is organized as follows. Below, we introduce notation and fix some fundamental definitions used later on. In the next section, we formulate the dissipative problems we wish to solve. In section 3, we present the discrete Riemannian gradient (DRG) methods, a convergence proof for the family of optimization methods obtained by applying DRG methods to Riemannian gradient flow problems, the Itoh–Abe discrete gradient generalized to manifolds, and the optimization algorithm obtained by applying the Itoh–Abe DRG to the gradient flow problem. In section 4, we provide numerical experiments to illustrate the use of DRGs in optimization, and in the final section we present conclusions and avenues for future work.

TABLE 1.1
Notational conventions.

Notation	Description
M	n-dimensional Riemannian manifold
T_pM	tangent space at $p \in M$ with zero vector 0_p
T_p^*M	cotangent space at $p \in M$
TM	tangent bundle of M
T^*M	cotangent bundle of M
$\mathfrak{X}(M)$	space of vector fields on M
$g(\cdot, \cdot)$	Riemannian metric on M
$\|\cdot\|_p$	norm induced on T_pM by g
$\{E_l\}_{l=1}^n$	g-orthogonal basis of T_pM

Notation and preliminaries. Some notation and definitions used in the following are summarized below and in Table 1.1. For a more thorough introduction to the concepts, see, e.g., [19] or [20].

On any differentiable manifold there is a duality pairing $\langle \cdot, \cdot \rangle : T^*M \times TM \to \mathbb{R}$ which we will denote as $\langle \omega, v \rangle = \omega(v)$. Furthermore, the Riemannian metric sets up an isomorphism between TM and T^*M via the linear map $v \mapsto g(v, \cdot)$. This map and its inverse, termed the musical isomorphisms, are known as the flat map $\flat : TM \to T^*M$ and sharp map $\sharp : T^*M \to TM$, respectively. The applications of these maps are also termed index raising and lowering when considering the tensorial representation of the Riemannian metric. Note that with the above notation we have the idiom $x^\flat(y) = \langle x^\flat, y \rangle = g(x, y)$.

On a Riemannian manifold, one can define gradients: For $V \in C^\infty(M)$, the (Riemannian) gradient with respect to g, $\operatorname{grad}_g V \in \mathfrak{X}(M)$, is the unique vector field such that $g(\operatorname{grad}_g V, X) = \langle \mathrm{d}V, X \rangle$ for all $X \in \mathfrak{X}(M)$. In the language of musical isomorphisms, $\operatorname{grad}_g V = (\mathrm{d}V)^\sharp$. For the remainder of this article, we will write $\operatorname{grad}V$ for the gradient and assume that it is clear from the context which g is to be used.

Furthermore, the *geodesic* between p and q is the unique curve of minimal length between p and q, providing a distance function $d_M : M \times M \to \mathbb{R}$. The geodesic γ passing through p with tangent v is given by the Riemannian exponential at p, $\gamma(t) = \exp_p(tv)$. For any p, $\exp_p$ is a diffeomorphism on a neighborhood N_p of 0_p, The image $\exp_p(S_p)$ of any star-shaped subset $S_p \subset N_p$ is called a normal neighborhood of p, and on this, $\exp_p$ is a radial isometry, i.e., $d_M(p, \exp_p(v)) = \|v\|_p$ for all $v \in S_p$.

2. The problem. We will consider ODEs of the form

$$(2.1) \qquad \dot{u} = F(u), \quad u(0) = u^0 \in M,$$

where $F \in \mathfrak{X}(M)$ has an associated energy $V : M \to \mathbb{R}$ dissipating along solutions of (2.1). That is, with $u(t)$ a solution of (2.1),

$$\frac{\mathrm{d}}{\mathrm{d}t} V(u) = \langle \mathrm{d}V(u), \dot{u} \rangle = \langle \mathrm{d}V(u), F(u) \rangle = g(\operatorname{grad}V(u), F(u)) \leq 0.$$

An example of such an ODE is the gradient flow. Given an energy V, the gradient flow of V with respect to a Riemannian metric g is

$$(2.2) \qquad \dot{u} = -\operatorname{grad}V(u),$$

which is dissipative since if $u(t)$ solves (2.2), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}V(u) = -g\left(\mathrm{grad}V(u), \mathrm{grad}V(u)\right) \leq 0.$$

Remark. This setting can be generalized by an approach similar to [21]. Suppose there exists a $(0,2)$ tensor field h on M such that $h(x,x) \leq 0$. We can associate to h the $(1,1)$ tensor field $H : TM \to TM$ given by $Hx = h(x, \cdot)^\sharp$. Consider the system

$$(2.3) \qquad\qquad\qquad \dot{u} = H\mathrm{grad}V(u).$$

This system dissipates V, since

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}V(u) &= \langle \mathrm{d}V(u), H\mathrm{grad}V(u)\rangle \\
&= g\left(\mathrm{grad}V(u), H\mathrm{grad}V(u)\right) \\
&= h\left(\mathrm{grad}V(u), \mathrm{grad}V(u)\right) \leq 0.
\end{aligned}$$

Any dissipative system of the form (2.1) can be written in this form on the set $M \backslash \{p \in M : g(F(p), \mathrm{grad}V(p)) = 0\}$ since, given F and V, we can construct h as follows:

$$h = \frac{1}{g(F, \mathrm{grad}V)} F^\flat \otimes F^\flat.$$

If $F = -\mathrm{grad}V$, we take $h = -g$ such that H becomes $-\mathrm{Id}$ and recover (2.2). In the following, we mainly discuss the case $F = -\mathrm{grad}V$ for the sake of notational clarity.

3. Numerical scheme. The discrete differentials in [7] are formulated such that they may be used on non-Riemannian manifolds. Since we restrict ourselves to Riemannian manifolds, we define their analogues: discrete Riemannian gradients. As with the discrete differentials, we shall make use of retractions as defined in [26].

DEFINITION 3.1. *Let $\phi : TM \to M$ and denote by ϕ_p the restriction of ϕ to T_pM. Then, ϕ is a retraction if the following conditions are satisfied:*
- *ϕ_p is smooth and defined in an open ball $B_{r_p(0_p)}$ of radius r_p around 0_p, the zero vector in T_pM.*
- *$\phi_p(v) = p$ if and only if $v = 0_p$.*
- *Identifying $T_{0_p}T_pM \simeq T_pM$, ϕ_p satisfies*

$$d\phi_p\big|_{0_p} = \mathrm{id}_{T_pM},$$

where id_{T_pM} denotes the identity mapping on T_pM.

From the inverse function theorem it follows that for any p, there exists a neighborhood $U_{p,\phi} \in T_pM$ of 0_p such that $\phi_p : U_{p,\phi} \to \phi_p(U_{p,\phi})$ is a diffeomorphism. In general, ϕ_p is not a diffeomorphism on the entirety of T_pM and so all the following schemes must be considered local in nature. The canonical retraction on a Riemannian manifold is the Riemannian exponential. It may be computationally expensive to evaluate even if closed expressions for geodesics are known, and so one often wishes to come up with less costly retractions if possible. We are now ready to introduce the notion of DRGs.

DEFINITION 3.2. *Given a retraction ϕ, a function $c : M \times M \to M$ where $c(p,p) = p$ for all $p \in M$, and a continuous $V : M \to \mathbb{R}$, then $\overline{\mathrm{grad}}V : M \times M \to TM$ is a DRG of V if it is continuous and, for all $p, q \in U_{c(p,q),\phi}$,*

$$(3.1) \qquad V(q) - V(p) = g\left(\overline{\mathrm{grad}}V(p,q), \phi_{c(p,q)}^{-1}(q) - \phi_{c(p,q)}^{-1}(p)\right),$$

$$(3.2) \qquad \overline{\mathrm{grad}}V(p,p) = \mathrm{grad}V|_p.$$

We formulate a numerical scheme for (2.2) based on this definition. Given times $0 = t_0 < t_1 < \ldots$, let u^k denote the approximation to $u(t_k)$ and let $\tau_k = t_{k+1} - t_k$. Then, we take

$$(3.3) \qquad u^{k+1} = \phi_{c^k}\left(W\left(u^k, u^{k+1}\right)\right),$$

$$(3.4) \qquad W(u^k, u^{k+1}) = \phi_{c^k}^{-1}(u^k) - \tau_k \overline{\mathrm{grad}}V(u^k, u^{k+1}),$$

where $c^k = c(u^k, u^{k+1})$. In the above and all of the following, we assume that u^k and u^{k+1} lie in $U_{c^k,\phi} \cap S_{c^k}$. The following proposition verifies that the scheme is dissipative.

PROPOSITION 3.3. *The sequence $\{u^k\}_{k\in\mathbb{N}}$ generated by the DRG scheme (3.3)–(3.4) satisfies $V(u^{k+1}) - V(u^k) \leq 0$ for all $k \in \mathbb{N}$.*

Proof. Using property (3.1) and (3.3) and (3.4), we get

$$
\begin{aligned}
V(u^{k+1}) - V(u^k) &= g\left(\overline{\mathrm{grad}}V\left(u^k, u^{k+1}\right), \phi_{c^k}^{-1}\left(u^{k+1}\right) - \phi_{c^k}^{-1}\left(u^k\right)\right) \\
&= g\left(\overline{\mathrm{grad}}V\left(u^k, u^{k+1}\right), W\left(u^k, u^{k+1}\right) - \phi_{c^k}^{-1}\left(u^k\right)\right) \\
&= -\tau_k g\left(\overline{\mathrm{grad}}V\left(u^k, u^{k+1}\right), \overline{\mathrm{grad}}V\left(u^k, u^{k+1}\right)\right) \leq 0. \qquad \square
\end{aligned}
$$

Remark. This extends naturally to schemes for (2.3) by exchanging (3.4) for

$$W(u^k, u^{k+1}) = \phi_{c^k}^{-1}(u^k) + \tau_k \overline{H}_{(u^k, u^{k+1})} \overline{\mathrm{grad}}V(u^k, u^{k+1}),$$

where $\overline{H}_{(p,q)}$ is the (1,1) tensor associated with a negative semidefinite (0,2) tensor field $\overline{h}_{(p,q)} : T_{c(p,q)}M \times T_{c(p,q)}M \to \mathbb{R}$ approximating $h|_p$ consistently.

Two DRGs, the average vector field DRG and the Gonzalez DRG, can be easily found by index raising the discrete differentials defined in [7]. We will later generalize the Itoh–Abe discrete gradient, but first we present a proof that the DRG scheme converges to a stationary point when used as an optimization algorithm. We will need the following definition of coercivity.

DEFINITION 3.4. *A function $V : M \to \mathbb{R}$ is coercive if, for all $v \in M$, every sequence $\{u^k\}_{k\in\mathbb{N}} \subset M$ such that $\lim_{k\to\infty} d_M(u^k, v) = \infty$ also satisfies $\lim_{k\to\infty} V(u^k) = \infty$.*

We will also need the following theorem from [28], concerning the boundedness of the sublevel sets $M_\mu = \{u \in M : V(u) \leq \mu\}$ of V.

THEOREM 3.5. *Assume M is unbounded. Then the sublevel sets of $V : M \to \mathbb{R}$ are bounded if and only if V is coercive.*

Proof. See [28, Chapter 1, Theorem 8.6] and the remarks below it. $\square$

Equipped with this, we present the following theorem, the proof of which is inspired by that of the convergence theorem in [11].

THEOREM 3.6. *Assume that M is geodesically complete, that $V : M \to R$ is coercive, bounded from below, and continuously differentiable, and that $\overline{\mathrm{grad}}V$ is continuous. Then, the iterates $\{u^k\}_{k\in\mathbb{N}}$ produced by applying the DRG scheme (3.3)–(3.4) with time steps $0 < \tau_{min} \leq \tau_k \leq \tau_{max}$ and $c^k = u^k$ or $c^k = u^{k+1}$ to the gradient flow of V satisfy*

$$\lim_{k\to\infty} \overline{\mathrm{grad}}V(u^k, u^{k+1}) = \lim_{k\to\infty} \mathrm{grad}V(u^k) = 0.$$

Additionally, there exists at least one accumulation point u^* of $\{u^k\}_{k\in\mathbb{N}}$, and any such accumulation point satisfies $\mathrm{grad}V(u^*) = 0$.

Proof. Since V is bounded from below and by Proposition 3.3, we have

$$C \leq V(u^{k+1}) \leq V(u^k) \leq \cdots \leq V(u^0)$$

such that, by the monotone convergence theorem, $V^* := \lim_{k \to \infty} V(u^k)$ exists. Furthermore, by property (3.1) and using the scheme (3.3)–(3.4),

$$\frac{1}{\tau_k} \left\| \phi_{c^k}^{-1}(u^k) - \phi_{c^k}^{-1}(u^{k+1}) \right\|_{c^k}^2 = \tau_k \left\| \overline{\mathrm{grad}} V(u^k, u^{k+1}) \right\|_{c^k}^2$$
$$= g\left(\overline{\mathrm{grad}} V(u^k, u^{k+1}), \phi_{c^k}^{-1}(u^k) - \phi_{c^k}^{-1}(u^{k+1}) \right)$$
$$= V(u^k) - V(u^{k+1}).$$

From this, it is clear that for any $i, j \in \mathbb{N}$,

$$\sum_{k=i}^{j-1} \tau_k \left\| \overline{\mathrm{grad}} V(u^k, u^{k+1}) \right\|_{c^k}^2 = V(u^i) - V(u^j) \leq V(u^0) - V^*$$

and

$$\sum_{k=i}^{j-1} \frac{1}{\tau_k} \left\| \phi_{c^k}^{-1}(u^k) - \phi_{c^k}^{-1}(u^{k+1}) \right\|_{c^k}^2 = V(u^i) - V(u^j) \leq V(u^0) - V^*.$$

In particular,

$$\sum_{k=0}^{\infty} \left\| \overline{\mathrm{grad}} V(u^k, u^{k+1}) \right\|_{c^k}^2 \leq \frac{V(u^0) - V^*}{\tau_{min}}$$

and

$$\sum_{k=0}^{\infty} \left\| \phi_{c^k}^{-1}(u^k) - \phi_{c^k}^{-1}(u^{k+1}) \right\|_{c^k}^2 \leq \tau_{max} \left(V(u^0) - V^* \right),$$

meaning

$$\lim_{k \to \infty} \left\| \overline{\mathrm{grad}} V(u^k, u^{k+1}) \right\|_{c^k} = 0,$$
$$\lim_{k \to \infty} \left\| \phi_{c^k}^{-1}(u^k) - \phi_{c^k}^{-1}(u^{k+1}) \right\|_{c^k} = 0.$$

Since u^{k+1} is in a normal neighborhood of c^k,

(3.5) $\qquad d_M(c^k, u^{k+1}) = d_M(c^k, \exp_{c^k}(\exp_{c^k}^{-1}(u^{k+1}))) = \| \exp_{c^k}^{-1}(u^{k+1}) \|_{c^k}.$

Introduce $\psi_{c^k} : T_{c^k}M \to T_{c^k}M$ by $\psi_{c^k} = \exp_{c^k}^{-1} \circ \phi_{c^k}$. Since both exp and ϕ are retractions,

$$\psi_{c^k}(0_{c^k}) = 0_{c^k},$$
$$D\psi_{c^k}|_{0_{c^k}} = \mathrm{id}_{T_{c^k}M}.$$

Thus, per the definition of Fréchet derivatives,

$$\psi_{c^k}(x) - \psi_{c^k}(0_{c^k}) - D\psi_{c^k}|_{0_{c^k}} x = \psi_{c^k}(x) - x = o(x);$$

in particular, choosing $x = \phi_{c^k}^{-1}(u^{k+1})$ we get

$$\exp_{c^k}^{-1}(u^{k+1}) - \phi_{c^k}^{-1}(u^{k+1}) = o(\| \phi_{c^k}^{-1}(u^{k+1}) \|_{c^k}),$$

meaning

$$(3.6) \qquad \| \exp_{c^k}^{-1}(u^{k+1}) \|_{c^k} \leq \| \phi_{c^k}^{-1}(u^{k+1}) \|_{c^k} + o(\| \phi_{c^k}^{-1}(u^{k+1}) \|_{c^k}).$$

Taking $c^k = u^k$ and combining (3.5) and (3.6) we find

$$d(u^k, u^{k+1}) = \| \exp_{c^k}^{-1}(u^{k+1}) \|_{c^k} \leq \| \phi_{c^k}^{-1}(u^{k+1}) \|_{c^k} + o(\| \phi_{c^k}^{-1}(u^{k+1}) \|_{c^k}).$$

Hence, since $\| \phi_{c^k}^{-1}(u^k) - \phi_{c^k}^{-1}(u^{k+1}) \|_{c^k} = \| \phi_{c^k}^{-1}(u^{k+1}) \|_{c^k}$ when $c^k = u^k$,

$$(3.7) \qquad \lim_{k \to \infty} d(u^k, u^{k+1}) \leq \lim_{k \to \infty} \left\| \phi_{c^k}^{-1}(u^k) - \phi_{c^k}^{-1}(u^{k+1}) \right\|_{c^k} = 0.$$

Note that we can exchange the roles of u^k and u^{k+1} and obtain the same result.

Since V is bounded from below, the sublevel sets M_μ of V are the preimages of the closed subsets $[C, \mu]$ and are hence closed as well. Since V is assumed to be coercive, by Theorem 3.5 the M_μ are bounded, and so since M is geodesically complete, by the Hopf–Rinow theorem the M_μ are compact [28]. In particular, $M_{V(u^0)}$ is compact such that $\overline{\text{grad}}V$ is uniformly continuous on $M_{V(u^0)} \times M_{V(u^0)}$ by the Heine–Cantor theorem. This means that for any $\epsilon > 0$ there exists $\delta > 0$ such that if $d_{M \times M}((u^k, u^{k+1}), (u^k, u^k)) = d_M(u^k, u^{k+1}) < \delta$, then

$$\left\| \overline{\text{grad}}V(u^k, u^{k+1}) - \text{grad}V(u^k) \right\|_{c^k} = \left\| \overline{\text{grad}}V(u^k, u^{k+1}) - \overline{\text{grad}}V(u^k, u^k) \right\|_{c^k} < \epsilon.$$

Since $d_M(u^k, u^{k+1}) \to 0$, given $\epsilon > 0$ there exists K such that for all $k > K$,

$$\left\| \text{grad}V(u^k) \right\|_{c^k} \leq \left\| \overline{\text{grad}}V(u^k, u^{k+1}) - \text{grad}V(u^k) \right\|_{c^k} + \left\| \overline{\text{grad}}V(u^k, u^{k+1}) \right\|_{c^k} \leq 2\epsilon.$$

This means

$$\lim_{k \to \infty} \text{grad}V(u^k) = 0.$$

Since $M_{V(u^0)}$ is compact, there exists a convergent subsequence $\{u^{k_l}\}$ with limit u^*. Since V is continuously differentiable,

$$\text{grad}V(u^*) = \lim_{l \to \infty} \text{grad}V(u^{k_l}) = 0. \qquad \qquad \square$$

Remark. In the above proof, we assumed $c^k = u^k$ or $c^k = u^{k+1}$. Although these choices may be desirable for practical purposes, as discussed in the next subsection, one can also make a more general choice. Specifically, if $\phi = \exp$ and c^k, let $\gamma^k(t)$ be the geodesic between u^k and u^{k+1} such that

$$\gamma^k(t) = \exp_{u^k}(tv^k),$$

where $v^k = \exp_{u^k}^{-1}(u^{k+1})$. Then, taking $c^k = \gamma^k(s)$ for some $s \in [0, 1]$, uniqueness of geodesics implies that

$$\exp_{c^k}(t\dot{\gamma}^k(s)) = \exp_{u^k}((t + s)v^k).$$

Hence,

$$\exp_{c^k}^{-1}(u^k) = -s\dot{\gamma}^k(s), \qquad \exp_{c^k}^{-1}(u^{k+1}) = (1 - s)\dot{\gamma}^k(s),$$

and so, since geodesics are constant speed curves,

$$d(u^k, u^{k+1}) = \|v\|_{u^k} = \|\dot{\gamma}^k(s)\|_{c^k} = \| \exp_{c^k}^{-1}(u^k) - \exp_{c^k}^{-1}(u^{k+1}) \|_{c^k}.$$

This means that (3.7) holds in this case. No other arguments in Theorem 3.6 are affected.

3.1. Itoh–Abe discrete Riemannian gradient. The Itoh–Abe discrete gradient [18] can be generalized to Riemannian manifolds.

PROPOSITION 3.7. *Given a continuously differentiable energy* $V : M \to \mathbb{R}$ *and an orthogonal basis* $\{E_j\}_{j=1}^n$ *for* $T_{c(u,v)}M$ *such that*

$$\phi_c^{-1}(v) - \phi_c^{-1}(u) = \sum_{i=1}^n \alpha_i E_i,$$

define $\overline{\mathrm{grad}}_{\mathrm{IA}}V : M \times M \to T_{c(u,v)}M$ *by*

$$\overline{\mathrm{grad}}_{\mathrm{IA}}V(u,v) = \sum_{j=1}^n a_j E_j,$$

where

$$a_j = \begin{cases} \dfrac{V(w_j) - V(w_{j-1})}{\alpha_j}, & \alpha_j \neq 0, \\ g(\mathrm{grad}V(w_{j-1}), d\phi_c\big|_{\eta_{j-1}} E_j), & \alpha_j = 0, \end{cases}$$

$$w_j = \phi_c(\eta_j), \quad \eta_j = \phi_c^{-1}(u) + \sum_{i=1}^j \alpha_i E_i.$$

Then, $\overline{\mathrm{grad}}_{\mathrm{IA}}V$ *is a DRG.*

Proof. Continuity of $\overline{\mathrm{grad}}_{\mathrm{IA}}V$ can be seen from the smoothness of the local coordinate frame $\{E_j\}_{j=1}^n$ and from the continuity of the $a_j(\alpha_j)$,

$$\begin{aligned} \lim_{\alpha_j \to 0} a_j(\alpha_j) &= \lim_{\alpha_j \to 0} \frac{V\left(\phi_c\left(\eta_{j-1} + \alpha_j E_j\right)\right) - V\left(\phi_c\left(\eta_{j-1}\right)\right)}{\alpha_j} \\ &= \frac{\mathrm{d}}{\mathrm{d}\alpha_j}\bigg|_{\alpha_j=0} V\left(\phi_c\left(\eta_{j-1} + \alpha_j E_j\right)\right) \\ &= \left\langle \mathrm{d}V\left(\phi_c\left(\eta_{j-1}\right)\right), d\phi_c\big|_{\eta_{j-1}} E_j \right\rangle \\ &= g(\mathrm{grad}V(w_{j-1}), d\phi_c\big|_{\eta_{j-1}} E_j). \end{aligned}$$

Property (3.1) holds since

$$\begin{aligned} g\left(\overline{\mathrm{grad}}_{\mathrm{IA}}V(u,v), \phi_c^{-1}(v) - \phi_c^{-1}(u)\right) &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i a_j g(E_i, E_j) \\ &= \sum_{j=1}^n V(w_j) - V(w_{j-1}) \\ &= V(w_n) - V(w_0) \\ &= V(v) - V(u). \end{aligned}$$

Furthermore, (3.2) holds since when $v = u$, all $\alpha_j = 0$, and $c(u,v) = u$ so that

$$\overline{\mathrm{grad}}_{\mathrm{IA}}V(u,u) = \sum_{j=1}^n g(\mathrm{grad}V(u), E_j)E_j = \mathrm{grad}V(u). \qquad \square$$

The map $\overline{\mathrm{grad}}_{\mathrm{IA}}V$ is called the Itoh–Abe DRG. For the Itoh–Abe DRG to be a computationally viable option it is important to compute the α_i efficiently. Consider, for instance, the gradient flow system. Applying the Itoh–Abe DRG to this we get the scheme

$$u^{k+1} = \phi_{c^k}\left(W(u^k, u^{k+1})\right),$$
$$W(u^k, u^{k+1}) = \phi_{c^k}^{-1}(u^k) - \tau_k\overline{\mathrm{grad}}_{\mathrm{IA}}V(u^k, u^{k+1}),$$

meaning

$$\phi_{c^k}^{-1}(u^{k+1}) - \phi_{c^k}^{-1}(u^k) = -\tau_k\overline{\mathrm{grad}}_{\mathrm{IA}}V(u^k, u^{k+1}),$$

and in coordinates

$$\sum_{i=1}^{n}\alpha_i E_i = -\tau_k\sum_{j=1}^{n}\frac{V(w_j) - V(w_{j-1})}{\alpha_j}E_j,$$

so that the α_i are found by solving the n coupled equations

$$\alpha_i = -\tau_k\frac{V(w_i) - V(w_{i-1})}{\alpha_i}.$$

Note that these equations in general are fully implicit in the sense that they require knowledge of the endpoint u^{k+1} since the w_i are dependent on c^k. However, if we take $c^k = u^k$, there is no dependency on the endpoint and all the above equations become scalar, although one must solve them successively. For this choice of c^k we present, as Algorithm 3.1, a procedure for solving the gradient flow problem on a Riemannian manifold with Riemannian metric g using the Itoh–Abe DRG.

Algorithm 3.1. DRG-OPTIM.

Choose $tol > 0$ and $u^0 \in M$. Set $k = 0$.
repeat
  Choose τ_k and an orthogonal basis $\{E_i^k\}_{i=1}^{n}$ for $T_{u^k}M$
  $v_0^k = u^k$
  $w_0^k = \phi_{u^k}^{-1}(v_0^k)$
  for $j = 1, \ldots, n$ **do**
    Solve $\alpha_j^k = -\tau_k\left(V\left(\phi_{u^k}(w_{j-1}^k + \alpha_j^k E_j^k)\right) - V\left(v_{j-1}^k\right)\right)/\alpha_j^k$
    $w_j^k = w_{j-1}^k + \alpha_j^k E_j^k$
    $v_j^k = \phi_{u^k}(w_j^k)$
  end for
  $u^{k+1} = v_n^k$
  $k = k + 1$
until $\left(V(u^k) - V(u^{k-1})\right)/V(u^0) < tol$

There is a caveat to this algorithm in that the α_j^k should be easy to compute. For example, it is important that the E_j and ϕ are chosen such that the difference $V(\phi_{u^k}(w_{j-1}^k + \alpha_j^k E_j^k)) - V(v_{j-1}^k)$ is cheap to evaluate. In many cases, M has a natural interpretation as a submanifold of Euclidean space defined locally by constraints $g : \mathbb{R}^m \to \mathbb{R}^n$, $M = \{y \in U \subset \mathbb{R}^m : g(y) = 0\}$. Then, one may find $\{E_j\}_{j=1}^{n}$ as an orthogonal basis for $\ker g'(c)$ and define ϕ_c implicitly by taking $q = \phi_c(v)$ such

that $q - (c + v) \in (T_c M)^\perp$ and $g(q) = 0$, as detailed in [6]. This requires the solution of a nonlinear system of equations for every coordinate update, which is computationally demanding compared to evaluating explicit expressions for $\{E_j\}_{j=1}^n$ and ϕ_c as is possible in special cases, such as those considered in section 4. To compute the α_j^k at each coordinate step one can use any suitable root finder, yet to stay in line with the derivative-free nature of Algorithm 3.1, one may wish to use a solver like the Brent–Dekker algorithm [3]. Also worth noting is that the parallelization procedure used in [22] works for Algorithm 3.1 as well.

4. Numerical experiments. This section concerns four applications of DRG methods to gradient flow systems. In each case, we specify all details needed to implement Algorithm 3.1: the manifold M, the retraction ϕ, and the basis vectors $\{E_k\}$. The first two examples are eigenvalue problems, included to illuminate implementational issues with examples in a familiar setting. We do not claim that our algorithm is competitive with other eigenvalue solvers but include these examples for the sake of exposition and to have problems with readily available reference solutions. The first of these is a simple Rayleigh quotient minimization problem, where issues of computational efficiency are raised. The second one concerns the Brockett flow on $SO(m)$, the space of orthogonal $m \times m$ matrices with unit determinant, and serves as an example of optimization on a Lie group. The remaining two problems are examples of manifold-valued image analysis problems concerning InSAR imaging and DTI, respectively. Specifically, the problems concern TV denoising of images obtained through these techniques [30]. The experiments do not consider the quality of the solution paths, i.e., numerical accuracy. For experiments of this kind, we refer to [5].

All programs used in the following were implemented as MATLAB functions, with critical functions implemented in C using the MATLAB EXecutable (MEX) interface when necessary. The code was executed using MATLAB (2017a release) running on a mid-2014 MacBook Pro with a four-core 2.5 GHz Intel Core i7 processor and 16 GB of 1600 MHz DDR3 RAM. We used a C language port of the built-in MATLAB function `fzero` for the Brent–Dekker algorithm implementation.

4.1. Eigenvalue problems. As an expository example, our first problem consists of finding the smallest eigenvalue/vector pair of a symmetric $m \times m$ matrix A by minimizing its Rayleigh quotient. We shall solve this problem using both the extrinsic and intrinsic view of the $(m-1)$-sphere. In the second example we consider the different approach to the eigenvalue problem proposed by Brockett in [4]. Here, the gradient flow on $SO(m)$ produces a diagonalizing matrix for a given symmetric matrix.

4.1.1. Eigenvalues via Rayleigh quotient minimization. In our first example, we wish to compute the smallest eigenvalue of a symmetric matrix $A \in \mathbb{R}^{m \times m}$ by minimizing the Rayleigh quotient

$$V(u) = u^T A u$$

with u on the $(m-1)$-sphere S^{m-1}.

Taking the extrinsic view, we regard S^{m-1} as a submanifold in $\mathbb{R}^m$, equipped with the standard Euclidian metric $g(x, y) = x^T y$. In this representation, $T_u S^{m-1}$ is the hyperplane tangent to u, i.e., $T_u S^{m-1} = \{x \in \mathbb{R}^m : x^T u = 0\}$. A natural choice of retraction is

$$\phi_p(x) = \frac{p + x}{\|p + x\|}.$$

There is a difficulty with this ϕ; it does not preserve sparsity, meaning Algorithm 3.1 will be inefficient as discussed above. To see this, consider that at each time step, to find the α_j^k, we must compute the difference

$$V\left(z_j^k\right) - V\left(z_{j-1}^k\right) = \left(z_j^k\right)^T A z_j^k - \left(z_{j-1}^k\right)^T A z_{j-1}^k$$

for some $z_{j-1}^k, z_j^k \in S^{m-1}$. We can compute this efficiently if $z_j^k = z_{j-1}^k + \delta$, where δ is sparse. Then,

$$V\left(z_j^k\right) - V\left(z_{j-1}^k\right) = 2\left(z_{j-1}^k\right)^T A\delta + \delta^T A\delta,$$

which is efficient since one may assume Az_{j-1}^k to be precomputed so that the computational cost is limited by the sparsity of δ. In our case, we have

$$z_{j-1}^k = \phi_c\left(w_{j-1}^k\right), \qquad z_j^k = \phi_c\left(w_{j-1}^k + \alpha_j^k E_j\right).$$

However, with ϕ_c as above, $\delta = \phi_c(w_{j-1}^k + \alpha_j^k E_j) - \phi_c(w_{j-1}^k)$ is nonsparse, and so computing the energy difference is costly.

Next, let us consider the intrinsic view of S^{m-1}, representing it in spherical coordinates $\theta \in \mathbb{R}^{m-1}$ by

$$u_1(\theta) = \cos(\theta_1),$$

$$u_r(\theta) = \cos(\theta_r)\prod_{i=1}^{r-1}\sin(\theta_i), \quad 1 < r < m,$$

$$u_m(\theta) = \prod_{i=1}^{m-1}\sin(\theta_i).$$

Due to the simple structure of $\mathbb{R}^{m-1}$, we take $\phi_\theta(\eta) = \theta + \eta$. Then, we have

$$u_r(\phi_\theta(\alpha E_l)) = u_r(\theta + \alpha E_l) = \begin{cases} u_r(\theta), & r < l, \\ \dfrac{\cos(\theta_l + \alpha)}{\cos(\theta_l)}u_r(\theta), & r = l, \\ \dfrac{\sin(\theta_l + \alpha)}{\sin(\theta_l)}u_r(\theta), & r > l. \end{cases}$$

Using this relation, the energy difference after a coordinate update becomes

$$V(u(\theta + \alpha E_l)) - V(u(\theta)) = 2\kappa_{1l}\sum_{i=1}^{l-1}u_i(\theta)u_l(\theta)A_{il} + 2\kappa_{2l}\sum_{i=1}^{l-1}\sum_{j=l+1}^{m}u_i(\theta)u_j(\theta)A_{ij}$$

$$+ 2\kappa_{3l}\sum_{j=l+1}^{m}u_l(\theta)u_j(\theta)A_{lj} + \kappa_{4l}\sum_{i=l+1}^{m}\sum_{j=l+1}^{m}u_i(\theta)u_j(\theta)A_{ij}$$

$$+ \kappa_{5l}u_l(\theta)u_l(\theta)A_{ll}$$

with

$$\kappa_{1l} = c_l - 1, \quad \kappa_{2l} = s_l - 1, \quad \kappa_{3l} = s_lc_l - 1, \quad \kappa_{4l} = s_l^2 - 1, \quad \kappa_{5l} = c_l^2 - 1,$$

where

$$c_l = \frac{\cos(\theta_l + \alpha)}{\cos(\theta_l)}, \quad s_l = \frac{\sin(\theta_l + \alpha)}{\sin(\theta_l)}.$$

With prior knowledge of $V(u(\theta))$ (and thus the four partial sums in the difference), evaluating $V(u(\theta + \alpha E_l)) - V(u(\theta))$ amounts to five scalar multiplications and four scalar additions after evaluating the κ_i^l. With correct bookkeeping, new sums can be evaluated from previous sums after coordinate updates, reducing the computational complexity of the algorithm. Although not producing an algorithm competitive with standard eigenvalue solvers, this example demonstrates that the correct choice of coordinates is vital to reducing the computational complexity of the Itoh–Abe DRG method.

4.1.2. Eigenvalues via Brockett flow. Among other things, the article of Brockett [4] discusses how one may find the eigenvalues of a symmetric matrix A by solving the following gradient flow problem on $M = \mathrm{SO}(m)$:

$$(4.1) \qquad \dot{Q} = -Q(DQ^T AQ - Q^T AQD).$$

Here, D is a real diagonal matrix with nonrepeated entries. It can be shown that $\lim_{t\to\infty} Q = Q^*$, where $(Q^*)^T AQ^* = \Lambda$ is diagonal and hence contains the eigenvalues of A, ordered as the entries of D. Equation (4.1) is the gradient flow of the energy

$$(4.2) \qquad V(Q) = \mathrm{tr}(AQ^T DQ)$$

with respect to the trace metric on $\mathrm{SO}(m)$. One can check that $\mathrm{SO}(m)$ is a Lie group [29], with Lie algebra

$$\mathfrak{so}(m) = \{B \in \mathbb{R}^{m\times m} : B^T = -B\}.$$

Also, since $\mathrm{SO}(m)$ is a matrix Lie group, the exponential coincides with the matrix exponential. However, we may consider using some other function as a retraction, such as the Cayley transform $\phi : \mathfrak{so}(m) \to \mathrm{SO}(m)$ given by

$$\phi(B) = (I - B)^{-1}(I + B).$$

Figure 4.1 shows the results of numerical tests with constant time step $\tau_k = 0.1$ and $m = 20$. In the left-hand panel, the evolution of the diagonal values of $Q^k AQ^k$ compared to the spectrum of A is shown; it is apparent that the diagonal values
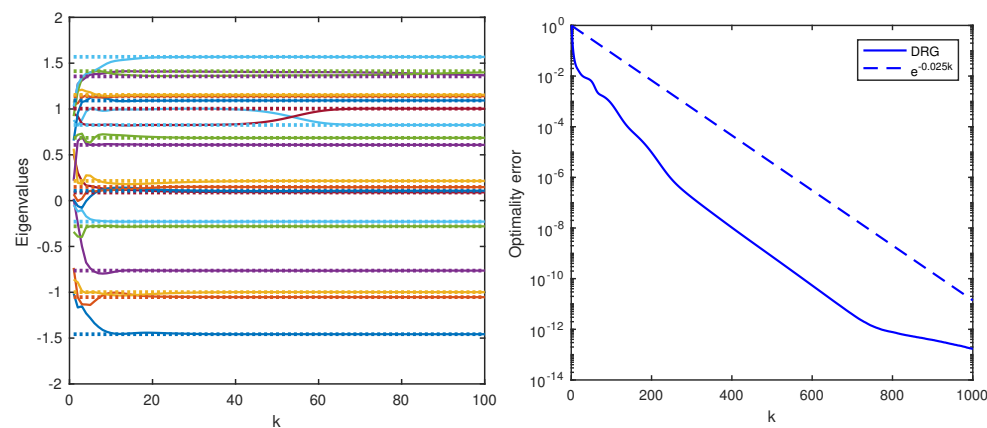

FIG. 4.1. *Brockett flow with $\tau_k = 0.1$ and 20 eigenvalues. Random initial matrix. Left: Evolution of eigenvalues. Right: Optimality error $(V(u^k) - V^*)/(V(u^0) - V^*)$.*

converge to the eigenvalues. The right-hand panel shows the convergence rate of Algorithm 3.1 to the minimal value V^* as computed with eigenvalues and eigenvectors from the MATLAB `eigen` function. It would appear that the convergence rate is linear, meaning $\|D - (Q^{k+1})^T A Q^{k+1}\| = C\|D - (Q^k)^T A Q^k\|$, with $C < 1$, which corresponds to an exponential reduction in $\|D - (Q^k)^T A Q^k\|$. No noteworthy difference was observed when using the matrix exponential in place of the Cayley transform.

4.2. Manifold valued imaging. In the following two examples we will consider problems from manifold valued two-dimensional imaging. We will in both cases work on a product manifold $\mathcal{M} = M^{l \times m}$ consisting of $l \times m$ copies of an underlying data manifold M. An element of M will in this case be called an *atom*, as opposed to the regular term *pixel*. As explained in [20], product manifolds of Riemannian manifolds are again Riemannian manifolds. The tangent spaces of product manifolds have a natural structure as direct sums, with $T_{(u_{11}, u_{12}, \ldots, u_{lm})}\mathcal{M} = \bigoplus_{i,j=1}^{l,m} T_{u_{ij}}M$, which induces a natural Riemannian metric $\mathcal{G} : T\mathcal{M} \times T\mathcal{M} \to \mathbb{R}$ fiberwise as

$$\mathcal{G}_{(u_{11}, u_{12}, \ldots, u_{lm})}((x_{11}, \ldots, x_{lm}), (y_{11}, \ldots, y_{lm})) = \sum_{i,j=1}^{l,m} g_{u_{ij}}(x_{ij}, y_{ij}).$$

Also, given a retraction $\phi : TM \to M$, one can define a retraction $\Phi : T\mathcal{M} \to \mathcal{M}$ fiberwise as

$$\Phi_{(u_{11}, u_{12}, \ldots, u_{lm})}(x_{11}, \ldots, x_{lm}) = (\phi_{u_{11}}(x_{11}), \phi_{u_{12}}(x_{12}), \ldots, \phi_{u_{lm}}(x_{lm})).$$

Discrete gradients were first used in optimization algorithms for image analysis in [11] and [22]. As an example of a manifold-valued imaging problem, consider TV denoising of manifold valued images [30], where one wishes to minimize, based on generalizations of the L^β and L^γ norms,

$$(4.3) \quad V(u) = \frac{1}{\beta} \sum_{i,j=1}^{l,m} \mathrm{d}(u_{ij}, s_{ij})^\beta + \lambda \left(\sum_{i,j=1}^{l-1,m} \mathrm{d}(u_{ij}, u_{i+1,j})^\gamma + \sum_{i,j=1}^{l,m-1} \mathrm{d}(u_{ij}, u_{i,j+1})^\gamma \right).$$

Here, $s = (s_{11}, \ldots, s_{lm}) \in \mathcal{M}$ is the input image, $u = (u_{11}, \ldots, u_{lm}) \in \mathcal{M}$ is the output image, λ is a regularization strength constant, and d is a metric on M, which we will take to be the geodesic distance induced by g.

4.2.1. InSAR image denoising. We first consider InSAR imaging, used in earth observation and terrain modeling [24]. In InSAR imaging, terrain elevation is measured by means of phase differences between laser pulses reflected from a surface at different times. Thus, the atoms g_{ij} are elements of $M = S^1$, represented by their phase angles: $-\pi < g_{ij} \leq \pi$. After processing, the phase data is *unwrapped* to form a single, continuous image of displacement data [9]. The natural distance function in this representation is the angular distance

$$\mathrm{d}(\varphi, \theta) = \begin{cases} |\varphi - \theta|, & |\varphi - \theta| \leq \pi, \\ 2\pi - |\varphi - \theta|, & |\varphi - \theta| > \pi. \end{cases}$$

Also, $T_\varphi M$ is simply $\mathbb{R}$, and ϕ is given, with $\underset{2\pi}{+}$ denoting addition modulo 2π, as

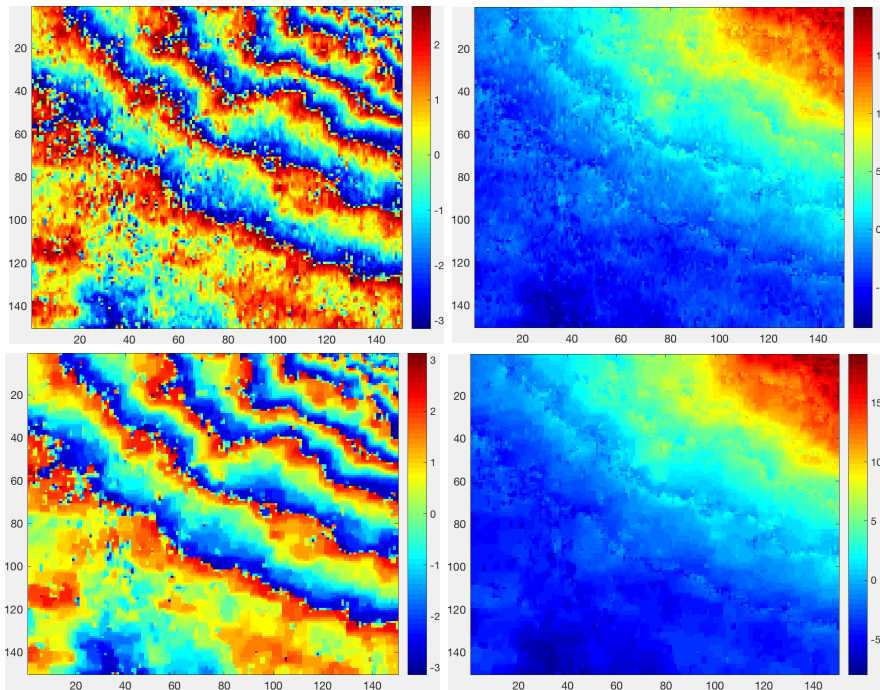$$\phi_\varphi(\theta_\varphi) = \left(\theta \underset{2\pi}{+} (\varphi + \pi) \right) - \pi.$$

FIG. 4.2. *Left column: Interferogram. Right column: Phase unwrapped image. Top row: Original image. Bottom row: L2 fidelity denoising, $\lambda = 0.3$.*
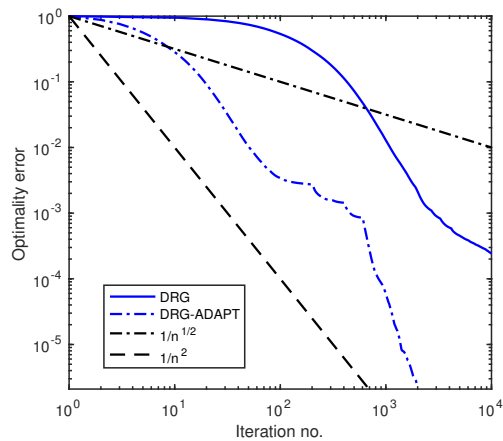

FIG. 4.3. *Logarithmic plot of optimality error $(V(u^k) - V^*)/(V(u^0) - V^*)$.*

Figure 4.2 shows the result of applying TV denoising to an InSAR image of a slope of Mt. Vesuvius, Italy, with $\beta = 2$. The left column shows the phase data, while the right-hand side shows the phase unwrapped data. The input image was taken from [23]. It is evident that the algorithm is successful in removing noise. Computation time was 0.1 seconds per iteration on a 150×150 image. A logarithmic plot showing convergence in terms of $(V(u^k) - V^*)/(V(u^0) - V^*)$ is shown in Figure 4.3, where V^* is a near-optimal value for V, obtained by iterating until $V(u^{k+1}) - V(u^k) \leq 10^{-15}$. The plot shows the behavior of Algorithm 3.1 with constant time steps $\tau_k = \tau_0 = 0.002$ and

an ad hoc adaptive method with $\tau_0 = 0.005$ where τ_k is halved each 200 iterations; for each of these strategies a separate V^* was found since they did not produce convergence to the same minimizer. The reason for the different minimizers is that the TV functional, and thus the minimization problem, is nonconvex in S^1 [27]. We can observe that the convergence speed varies between $\mathcal{O}(1/k)$ and $\mathcal{O}(1/k^2)$, with faster convergence for the ad hoc adaptive method. The reason for this sublinear convergence as compared to the linear convergence observed in the Brockett flow case may be the nonconvexity.

4.2.2. DTI image denoising. DTI is a medical imaging technique where the goal is to make spatial samples of the tensor specifying the diffusion rates of water in biological tissue. The tensor is assumed to be, at each point (i, j), represented by a matrix $A_{ij} \in \mathrm{Sym}^+(3)$, the space of 3×3 symmetric positive definite (SPD) matrices. Experimental measurements of DTI data are, as with other MRI techniques, contaminated by Rician noise [12], which one may attempt to remove by minimizing (4.3) with an appropriate choice of Riemannian structure on $\mathcal{M} = \mathrm{Sym}^+(3)^{m \times l}$.

As above, since the manifold we are working on is a product manifold, it suffices to define the Riemannian structure on $\mathrm{Sym}^+(3)$. First off, one should note that $T_A\mathrm{Sym}^+(3)$ can be identified with $\mathrm{Sym}(3)$, the space of symmetric 3×3 matrices [19]. In [30], the authors consider equipping $\mathrm{Sym}^+(3)$ with the affine invariant Riemannian metric given pointwise as

$$g_A(X, Y) = \mathrm{tr}\left(A^{-\frac{1}{2}} X A^{-1} Y A^{-\frac{1}{2}}\right),$$

and for purposes of comparison, so shall we. The space $\mathrm{Sym}^+(3)$ equipped with this metric is a Cartan–Hadamard manifold [19] and thus is complete, meaning that Theorem 3.6 holds. This metric induces the explicitly computable geodesic distance

$$d(A, B) = \sqrt{\sum_{i=1}^{3} \log(\kappa_i)^2}$$

on $\mathrm{Sym}^+(3)$, where κ_i are the eigenvalues of $A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$. Furthermore, the metric induces a Riemannian exponential given by

$$\exp_A(Y) = A^{1/2} e^{A^{-1/2} Y A^{-1/2}} A^{1/2},$$

where e denotes the matrix exponential, and $A^{1/2}$ is the matrix square root of A. We could choose the retraction as $\phi = \exp$, but there are less computationally expensive options that do not involve computing matrix exponentials. More specifically, we will make use of the second-order approximation of the exponential,

$$\phi_A(Y) = A + Y + \frac{1}{2} Y A^{-1} Y.$$

While a first-order expansion is also a retraction, there is no guarantee that $A + Y \in \mathrm{Sym}^+(3)$, whereas the second-order expansion, which can be written in the form

$$\phi_A(Y) = \frac{1}{2} A + \frac{1}{2} \left(A^{\frac{1}{2}} + A^{-\frac{1}{2}} Y\right)^T \left(A^{\frac{1}{2}} + A^{-\frac{1}{2}} Y\right),$$

is clearly SPD since A is so. Note that using a sparse basis E_{ij} (in our example we use $E_{ij} = e_i e_j^T + e_j e_i^T$) for the space $\mathrm{Sym}(3)$, evaluating $\phi_A(X + \alpha E_{ij})$ amounts to, at
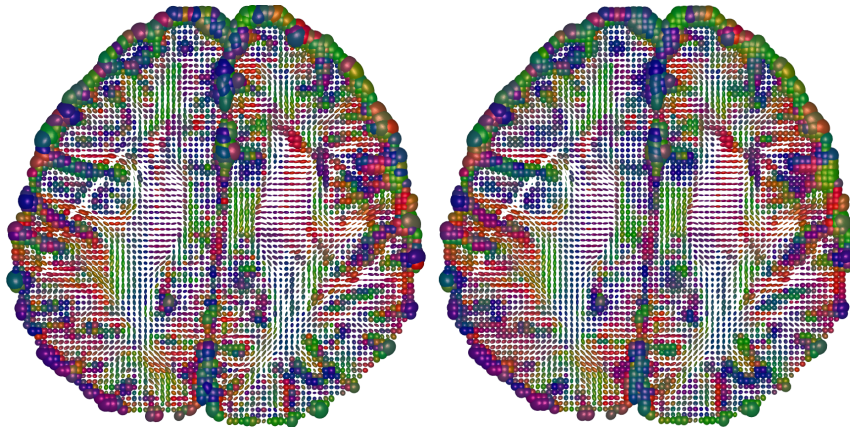
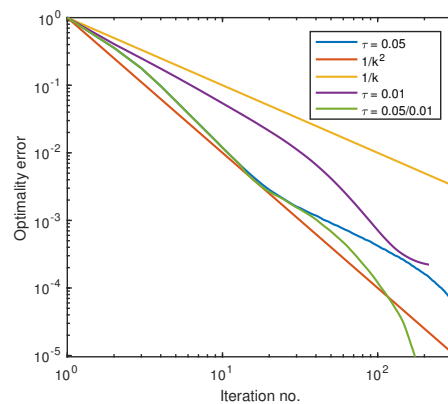FIG. 4.4. *DTI scan, axial slice. Left: Noisy image. Right: Denoised with $\beta = 2$, $\lambda = 0.05$.*

FIG. 4.5. *Logarithmic plot of optimality error.*

most, four scalar updates when $\phi_A(X)$ and A^{-1} is known, as is possible with proper
bookkeeping in the software implementation. Also, since all matrices involved are
3×3 SPD matrices, one may find eigenvalues and eigenvectors directly, thus allowing
for fast computations of matrix square roots and, consequently, geodesic distances.

Figure 4.4 shows an example of denoising DTI images using the TV regularizer.
The data is taken from the publicly available Camino data set [8]. The DTI tensor
has been calculated from underlying data using linear least-squares fitting and is
subject to Rician noise (left-hand side), which is mitigated by TV denoising (right-
hand side). The denoising procedure took about 7 seconds for 57 iterations, on a
72×73 image. The algorithm was stopped when the relative change in energy, $(V(u^0) -
V(u^k))/V(u^0)$, dropped below 10^{-5}. Each atom $A \in \mathrm{Sym}^+(3)$ is visualized by an
ellipsoid with the eigenvectors of A as principal semiaxes, scaled by the corresponding
eigenvalues. The colors are coded to correspond to the principal direction of the
major axis, with red denoting left-right orientation, green anterior-posterior, and blue
inferior-superior. Figure 4.5 shows the convergence behavior of Algorithm 3.1, with
three different time steps: $\tau = 0.05$, $\tau = 0.01$, and a mixed strategy of using $\tau = 0.05$
for 12 steps, then changing to $\tau = 0.01$. Also, baseline rates of $1/k^2$ and $1/k$ are shown.
It is apparent that the choice of time step has a great impact on the convergence rate

and that simply changing the time step from $\tau = 0.05$ to $\tau = 0.01$ is effective in speeding up convergence. This would suggest that time step adaptivity is a promising route for acceleration of these methods.

5. Conclusion and outlook. We have extended discrete gradient methods to Riemannian manifolds and shown how they may be applied to gradient flows. The Itoh–Abe discrete gradient has been formulated in a manifold setting; this is, to the best of our knowledge, the first time this has been done. In particular, we have used the Itoh–Abe DRG on gradient systems to produce a derivative-free optimization algorithm on Riemannian manifolds. This optimization algorithm has been proven to converge under reasonable conditions and shows promise when applied to the problem of denoising manifold valued images using the TV approach of [30].

As with the algorithm in the Euclidian case, there are open questions. The first question is which convergence rate estimates can be made; one should especially consider the linear convergence exhibited in the Brockett flow problem, and the rate observed in Figure 4.5 which approaches $1/k^2$. A second question is how to formulate a rule for choosing step sizes so as to accelerate convergence toward minimizers. There is also the question of how the DRG methods perform as ODE solvers for dissipative problems on Riemannian manifolds, in particular, convergence properties, stability, and convergence order. The above discussion is geared toward optimization applications due to the availability of optimization problems, but it would be of interest to see how the methods work as ODE solvers in their own right similar to the analysis and experiments done in [5].

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2009.

[2] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, *Newton's method on Riemannian manifolds and a geometric model for the human spine*, IMA J. Numer. Anal., 22 (2002), pp. 359–390.

[3] R. P. Brent, *An algorithm with guaranteed convergence for finding a zero of a function*, Comput. J., 14 (1971), pp. 422–425.

[4] R. W. Brockett, *Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems*, in Proceedings of the 27th IEEE Conference on Decision and Control, IEEE, 1988, pp. 799–803.

[5] E. Celledoni, S. Eidnes, B. Owren, and T. Ringholm, *Energy Preserving Methods on Riemannian Manifolds*, preprint, arXiv:1805.07578, 2018.

[6] E. Celledoni and B. Owren, *A class of intrinsic schemes for orthogonal integration*, SIAM J. Numer. Anal., 40 (2002), pp. 2069–2084.

[7] E. Celledoni and B. Owren, *Preserving first integrals with symmetric Lie group methods*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 977–990.

[8] P. Cook, Y. Bai, S. Nedjati-Gilani, K. Seunarine, M. Hall, G. Parker, and D. Alexander, *Camino: Open-source diffusion-MRI reconstruction and processing*, in Proceedings of ISMRM, Seattle, WA, 2006.

[9] R. M. Goldstein, H. A. Zebker, and C. L. Werner, *Satellite radar interferometry: Two-dimensional phase unwrapping*, Radio Sci., 23 (1988), pp. 713–720.

[10] O. Gonzalez, *Time integration and discrete Hamiltonian systems*, J. Nonlinear Sci., 6 (1996), pp. 449–467.

[11] V. Grimm, R. I. McLachlan, D. I. McLaren, G. Quispel, and C. Schönlieb, *Discrete gradient methods for solving variational image regularisation models*, J. Phys. A, 50 (2017), 295201.

[12] H. Gudbjartsson and S. Patz, *The Rician distribution of noisy MRI data*, Magn. Reson. Med., 34 (1995), pp. 910–914.

[13] E. Hairer and C. Lubich, *Energy-diminishing integration of gradient systems*, IMA J. Numer. Anal., 34 (2013), pp. 452–461.

[14] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Ser. Comput. Math. 31, Springer, New York, 2006.

[15] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations.* II, 2nd ed., Springer Ser. Comput. Math. 14, Springer, New York, 1996.

[16] A. Harten, P. D. Lax, and B. van Leer, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev., 25 (1983), pp. 35–61.

[17] A. Humphries and A. Stuart, *Runge–Kutta methods for dissipative and gradient dynamical systems*, SIAM J. Numer. Anal., 31 (1994), pp. 1452–1485.

[18] T. Itoh and K. Abe, *Hamiltonian-conserving discrete canonical equations based on variational difference quotients*, J. Comput. Phys., 76 (1988), pp. 85–102.

[19] S. Lang, *Fundamentals of Differential Geometry*, Grad. Texts in Math. 191, Springer, New York, 2012.

[20] J. M. Lee, *Riemannian Manifolds: An Introduction to Curvature*, Grad. Texts in Math. 176, Springer, New York, 2006.

[21] R. I. McLachlan, G. R. W. Quispel, and N. Robidoux, *Geometric integration using discrete gradients*, Philos. Trans. A, 357 (1999), pp. 1021–1045.

[22] T. Ringholm, J. Lazić, and C.-B. Schönlieb, *Variational image regularization with Euler's elastica using a discrete gradient scheme*, SIAM J. Imaging Sci., to appear.

[23] F. Rocca, C. Prati, and A. Ferretti, *An overview of ERS-SAR interferometry*, in Proceedings of the ERS Symposium "Space in the Service of our Environment," 1997.

[24] P. A. Rosen, S. Hensley, I. R. Joughin, F. K. Li, S. N. Madsen, E. Rodriguez, and R. M. Goldstein, *Synthetic aperture radar interferometry*, Proc. IEEE, 88 (2000), pp. 333–382.

[25] S. Sato, T. Matsuo, H. Suzuki, and D. Furihata, *A Lyapunov-type theorem for dissipative numerical integrators with adaptive time-stepping*, SIAM J Numer. Anal., 53 (2015), pp. 2505–2518.

[26] M. Shub, *Some remarks on dynamical systems and numerical analysis*, in Proceedings of the 7th ELAM, 1986, pp. 69–92.

[27] E. Strekalovskiy and D. Cremers, *Total variation for cyclic structures: Convex relaxation and efficient minimization*, in Proceedings of CVPR, IEEE, 2011, pp. 1905–1911.

[28] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Math. Appl. 297, Springer, New York, 1994.

[29] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Grad. Texts in Math. 94, Springer, New York, 2013.

[30] A. Weinmann, L. Demaret, and M. Storath, *Total variation regularization for manifold-valued data*, SIAM J. Imaging Sci., 7 (2014), pp. 2226–2257.