

A STUDY OF THE ASYMPTOTIC HOLONOMIC EFFICIENCY PROBLEM*

JIANGHAI HU[†] AND SLOBODAN SIMIC[‡]

Abstract. In this paper we study an asymptotic version of the holonomic efficiency problem originated in the study of swimming microorganism. Given a horizontal distribution on a vector bundle, the holonomy of a loop in the base space is the displacement along the fiber direction of the end points of its horizontal lift. The holonomic efficiency problem is to find the most efficient loop in the base space in terms of gaining holonomy, where the cost of the base loop is measured by a subriemannian metric, and the holonomy gained is compared using a test function. We introduce the notions of rank and asymptotic holonomy, and characterize them through the series expansions of holonomy as a function of the loop scale. In the rank two case we prove that for convex test functions the most efficient base loops are simple circles, and solve these loops for linear and norm test functions. In the higher rank case the analytical solutions are outlined for some special instances of the problem. An example of a turning linked-mass system is worked out in detail to illustrate the results.

Key words. Nonholonomic systems, optimal control, holonomy, sub-riemannian geometry.

AMS subject classifications. 53C17, 49K15, 93B18, 70F25, 70G45.

1. Introduction. The isoholonomic problem has applications in a variety of fields, for example, the falling cat problem [2] in mechanics, the swimming microorganism at low Reynolds number [3] in biology, and the Berry phase in quantum mechanics, etc. [4]. Generally speaking, for a loop c in the base space M of a principal bundle $\pi : Q \rightarrow M$ with a horizontal distribution \mathcal{H} , its holonomy is the vertical displacement of the end points of its horizontal lift in Q . The holonomic efficiency of c can then be defined as the ratio of a certain functional of the holonomy and some quantity such as the length or energy that characterizes the cost for traversing c . The isoholonomic problem tries to find loop c with the highest holonomic efficiency.

In the context of micro-swimming, various notions of holonomic efficiency have been proposed. To name a few, in [5] the efficiency of a swimming stroke is defined as the ratio of the square of average speed achieved by the stroke to the average power output required; this notion of efficiency is scaled by a characteristic thrust to obtain the dimensionless Froude's efficiency studied in [6]; while in [7] the efficiency is the ratio of the product of average speed and a characteristic thrust to the average power. Another notion of efficiency is proposed in [8] that is invariant to temporal and spatial rescaling. See [9] for more discussions on the various notions of efficiency in the microswimming problem, and [4, 10, 11] for applications in other areas. The isoholonomic problem can be formulated as a special class of optimal control problems, whose solution has been studied, for example, in [12, 13, 14, 15].

There are several limitations with these previous work that motivate the research in this paper. First of all, although all of them deal with the asymptotic case, a rigorous formulation of the asymptotic holonomic efficiency problem has not been adequately addressed yet. Secondly, the problems studied so far focus on the non-degenerate (rank two) case only, while the degenerate (higher rank) case has been

*PART OF THE RESULTS IN THIS PAPER HAS APPEARED IN [1].

[†]School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Ave., West Lafayette, IN, 47906, USA. E-mail: jianghai@purdue.edu

[‡]Department of Mathematics, San Jose State University, San Jose, CA 95192-0103, USA. Email: simic@math.sjsu.edu.

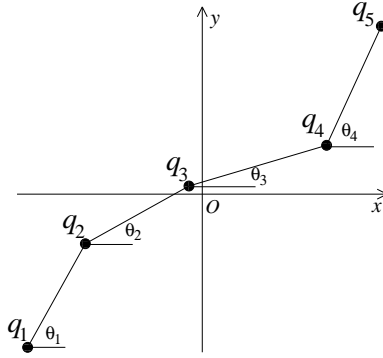


FIG. 1.1. A linked-mass system with five nodes and four segments.

largely ignored. Thirdly, the various notions of efficiency proposed in the literature, with the exception of [4], are defined through linear functionals (test functions) of the holonomy, whereas in some cases general test functions could be more desirable.

In this paper we propose a general framework to study the asymptotic holonomic efficiency problem. The concepts of rank and asymptotic holonomy are defined, and, using the notion of test functions, we propose two optimization problems that are dual to each other to study the most efficient way to gain asymptotic holonomy: the generalized isoholonomic and isoperimetric problems. Both the rank two case and the higher rank case are considered. In the rank two case, we extend the well known result that optimal solutions are circles from the linear test function case to the general convex test function case, and propose procedures for finding these circles for norm test functions. For the higher rank case, we focus on a special family of distributions with arbitrary rank, and find the optimal solution through optimal control theory.

For simplicity, our problems are formulated on the trivial vector bundle $\pi : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^n$. However, due to their asymptotic nature, our results can be easily extended to the case of more general state spaces such as principal bundles.

1.1. Motivating Example. We start by introducing a motivating example of a snake-like linked-mass system moving on a plane first reported in [1]. A relevant but more complicated model is the molecule model studied in [16].

The system consists of $n + 2$ unit point masses (nodes) connected subsequently by $n + 1$ rigid links (segments) of unit length and zero mass. Figure 1.1 shows an example with $n = 3$. Given a fixed coordinate system with the origin O , denote by $q_1(t), \dots, q_{n+2}(t) \in \mathbb{R}^2$ the locations of the $n + 2$ nodes at time $t \geq 0$. Assume without loss of generality that $\sum_{i=1}^{n+2} q_i(0) = 0$, i.e., the system is initially centered at the origin O . Suppose that the system is subject to no external forces. Then its total linear momentum and total angular momentum (using the origin O as the center of rotation) are conserved:

$$\sum_{i=1}^{n+2} \dot{q}_i \equiv 0, \quad (1.1)$$

$$\sum_{i=1}^{n+2} q_i \times \dot{q}_i \equiv 0. \quad (1.2)$$

Note that the zero-mass links do not contribute to the above computation.

Condition (1.1) then implies that $\sum_{i=1}^{n+2} q_i \equiv 0$. So the configuration of the system is uniquely determined by the angles $\theta_1, \dots, \theta_{n+1}$, where θ_i is the angle $q_{i+1} - q_i$ makes with the positive x -axis, $i = 1, \dots, n+1$. Each θ_i takes values in \mathbb{R} modulo 2π , namely, the 1-torus $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$; so $(\theta_1, \dots, \theta_{n+1})$ takes values in the $(n+1)$ -torus \mathbb{T}^{n+1} , which is the *configuration space* of the system.

Remark 1 *There is a natural bundle structure on \mathbb{T}^{n+1} . \mathbb{T} acts on \mathbb{T}^{n+1} by*

$$R_\theta(\theta_1, \dots, \theta_{n+1}) = (\theta_1 + \theta, \dots, \theta_{n+1} + \theta), \quad \forall \theta \in \mathbb{T}, (\theta_1, \dots, \theta_{n+1}) \in \mathbb{T}^{n+1}. \quad (1.3)$$

The effect of R_θ on any configuration is a rotation of θ counterclockwise. Each orbit of this action consists of configurations with the same shape but different orientations; and configurations in different orbits have different shapes. Thus the shape space of the linked-mass system can be defined as the set of all R -orbits in \mathbb{T}^{n+1} , i.e., $\mathbb{T}^{n+1}/\mathbb{T}$, which is an n -torus topologically. The quotient map $\pi : \mathbb{T}^{n+1} \rightarrow \mathbb{T}^{n+1}/\mathbb{T}$ defines \mathbb{T}^{n+1} as a \mathbb{T} -bundle over $\mathbb{T}^{n+1}/\mathbb{T}$, whose fibers are exactly the R -orbits.

For given $\theta_1, \dots, \theta_{n+1}$, the corresponding q_1, \dots, q_{n+2} satisfying $\sum_{i=1}^{n+2} q_i = 0$ are

$$q_1 = -\frac{1}{n+2} \sum_{j=1}^{n+1} (n+2-j)(\cos \theta_j, \sin \theta_j)^T, \quad (1.4)$$

$$q_i = q_1 + \sum_{j=1}^{i-1} (\cos \theta_j, \sin \theta_j)^T, \quad i = 2, \dots, n+2. \quad (1.5)$$

Equations (1.4) and (1.5) together define an embedding of the configuration space \mathbb{T}^{n+1} into \mathbb{R}^{2n+4} . Thus \mathbb{T}^{n+1} inherits isometrically via this embedding a riemannian metric $\langle \cdot, \cdot \rangle$ from the standard metric on \mathbb{R}^{2n+4} . From the computation in Appendix A, $\langle \cdot, \cdot \rangle$ can be determined as

$$g_{ij} \triangleq \left\langle \frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j} \right\rangle = \Delta_{ij} \cos(\theta_i - \theta_j), \quad 1 \leq i, j \leq n+1, \quad (1.6)$$

where Δ_{ij} are constants defined by

$$\Delta_{ij} = \begin{cases} \frac{i(n+2-j)}{n+2}, & \text{if } i < j, \\ \frac{(n+2-i)j}{n+2}, & \text{if } i \geq j. \end{cases}$$

Suppose that the trajectory of the linked-mass system over a time interval $I = [0, 1]$ is given by a curve γ in \mathbb{T}^{n+1} . Unless otherwise stated, we assume that all curves in this paper are defined on I . Define

$$L(\gamma) = \int_0^1 \|\dot{\gamma}\| dt, \quad E(\gamma) = \int_0^1 \|\dot{\gamma}\|^2 dt$$

as the *length* and the *energy* of γ , respectively, where $\|\cdot\|$ is the norm corresponding to $\langle \cdot, \cdot \rangle$. From the definition of $\langle \cdot, \cdot \rangle$, we have $L(\gamma) = \int_0^1 (\sum_{i=1}^{n+2} \|\dot{q}_i\|^2)^{1/2} dt$ and $E(\gamma) = \int_0^1 \sum_{i=1}^{n+2} \|\dot{q}_i\|^2 dt$, where q_1, \dots, q_{n+2} are the positions of the nodes corresponding to γ . A physical explanation of the expression of $E(\gamma)$ is that, since the links in the system have zero mass, their rotations and translations do not contribute to the total

energy; hence the energy of the path (maneuver) γ is the time integral of (twice) the total instantaneous kinetic energy of the nodes only, which is derived in Appendix A in the alternative coordinates $(\theta_1, \dots, \theta_{n+1})$.

With this metric on \mathbb{T}^{n+1} , we now study the geometric implication of the constraint (1.2). It can be verified that a curve $\gamma = (\theta_1, \dots, \theta_{n+1})$ in \mathbb{T}^{n+1} satisfies the constraint (1.2) if and only if

$$\sum_{i,j=1}^{n+1} \Delta_{ij} \cos(\theta_i - \theta_j) \dot{\theta}_j = 0, \quad (1.7)$$

or equivalently, if and only if $\Theta(\dot{\gamma}) = 0$, where Θ is a one-form on \mathbb{T}^{n+1} defined by

$$\Theta = \sum_{i,j=1}^{n+1} \Delta_{ij} \cos(\theta_i - \theta_j) d\theta_j. \quad (1.8)$$

In other words, γ must be a *horizontal* curve for the co-dimension one distribution $\mathcal{H} \triangleq \ker \Theta$ on \mathbb{T}^{n+1} . The restriction of $\langle \cdot, \cdot \rangle$ to \mathcal{H} defines a sub-riemannian metric $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. In this metric the sub-riemannian length of a horizontal curve γ is the same as its riemannian length $L(\gamma)$.

We observe that the horizontal distribution \mathcal{H} defined above from equation (1.7) is naturally induced by the metric $\langle \cdot, \cdot \rangle$ given in (1.6). In fact, at each point $\theta = (\theta_1, \dots, \theta_{n+1}) \in \mathbb{T}^{n+1}$, the codimensional-one horizontal space $\mathcal{H}_{\theta} = \ker \Theta_{\theta}$ can be easily verified as the orthogonal complement of the bundle direction of $\pi : \mathbb{T}^{n+1} \rightarrow \mathbb{T}^n$, namely, $(1, \dots, 1)$, under the metric $\langle \cdot, \cdot \rangle$:

$$\mathcal{H}_{\theta} = \{v \in T_{\theta} \mathbb{T}^{n+1} \simeq \mathbb{R}^{n+1} \mid \langle v, (1, \dots, 1) \rangle = 0\}.$$

This fact has been observed in the general setting of rotation and vibration motions of molecules in [16]. The connections resulting from such horizontal distributions are called the natural mechanical connections [17].

1.2. Objective and Overview of the Paper. We are interested in finding the most efficient way for the linked-mass system to turn. More precisely, among all the maneuvers γ that guide the system from a starting configuration $(\theta_1^0, \dots, \theta_{n+1}^0)$ at time 0 to a desired configuration $(\theta_1^0 + \theta, \dots, \theta_{n+1}^0 + \theta)$ at time 1 that has the same shape but a different orientation, subject to the constraint (1.2) of zero total angular momentum, we want to find the one (or ones) with minimal energy (or minimal length $L(\gamma)$, which are equivalent up to reparameterizations). In light of the above discussion, the solutions to this problem are the shortest horizontal curves in \mathbb{T}^{n+1} connecting $(\theta_1^0, \dots, \theta_{n+1}^0)$ to $(\theta_1^0 + \theta, \dots, \theta_{n+1}^0 + \theta)$, which are necessarily distance-minimizing sub-riemannian geodesics in $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

Remark 2 In definitions (1.6) and (1.8) the terms involving θ_i 's are of the form $\theta_i - \theta_j$, which remain unchanged under the action R . Thus both \mathcal{H} and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ are invariant along the fibers of $\pi : \mathbb{T}^{n+1} \rightarrow \mathbb{T}^{n+1}/\mathbb{T}$, and together they specify a sub-riemannian geometry invariant with respect to π (see Section 2). In this perspective, the problem under study is to determine the shortest horizontal curve connecting two points $(\theta_1, \dots, \theta_{n+1})$ and $R_{\theta}(\theta_1, \dots, \theta_{n+1})$ in the same fiber.

Unfortunately, solutions to the above formulated problem are usually impossible to obtain analytically due to its global nature (the starting and ending configurations

could be far away from each other). In this paper, we shall instead study an asymptotic (local) version of the problem, i.e., what is the most efficient way for the linked-mass system to turn if it can only exert an increasingly small amount of energy. The exact formulation of the asymptotic problem will be given in Section 2 in the more general context of co-dimension k distribution on \mathbb{R}^{n+k} . In particular, we define the notions of rank and asymptotic holonomy, and, using test functions, propose an optimization problem that generalizes the efficiency problems studied in the literature. In Section 3, we focus on the rank two case and prove that, for convex test functions, at least one of the solutions to the optimization problem is given by a simple circle contained in a two-dimensional plane. Although such solutions are well known in the literature when the test function is linear, our results hold for arbitrary convex test function. Detailed procedures are also outlined for finding the solution when the test function is a norm. The higher rank case, on the other hand, is much more complicated. In Section 4, we solve the problem for a special family of distributions with rank higher than two, and in Section 5 use the result to obtain the asymptotically most efficient maneuver for the linked-mass system in Section 1. Section 6 extends the results to connection on principal bundles.

2. Problem Formulation. We now formulate the problem in the general setting of co-dimension k distributions on the Euclidean space \mathbb{R}^{n+k} for some $n, k \geq 1$. The projection $\pi : (x_1, \dots, x_{n+k}) \in \mathbb{R}^{n+k} \mapsto (x_1, \dots, x_n) \in \mathbb{R}^n$ defines \mathbb{R}^{n+k} as a trivial vector bundle over \mathbb{R}^n whose fiber over each $m \in \mathbb{R}^n$ is given by $\pi^{-1}(m) \simeq \mathbb{R}^k$. We shall first review some relevant concepts in sub-riemannian geometry. A comprehensive introduction on this topic can be found in [10].

2.1. Co-dimension k distributions and sub-riemannian metrics on \mathbb{R}^{n+k} . Let $\Theta = (\Theta_1, \dots, \Theta_k)$ be an \mathbb{R}^k -valued one-form on \mathbb{R}^{n+k} with components

$$\Theta_j = dx_{n+j} - \sum_{i=1}^n \alpha_i^j dx_i, \quad 1 \leq j \leq k, \quad (2.1)$$

for some C^∞ functions $\alpha_i^j : \mathbb{R}^{n+k} \rightarrow \mathbb{R}$. Then $\mathcal{H} = \ker \Theta$ is a co-dimension k distribution on \mathbb{R}^{n+k} . The *horizontal space* \mathcal{H}_q at each $q \in \mathbb{R}^{n+k}$ is the kernel of Θ_q in $T_q \mathbb{R}^{n+k}$, which can be thought of as an n -dimensional subspace of \mathbb{R}^{n+k} , i.e., $\mathcal{H}_q = \{(v_1, \dots, v_{n+k}) \in \mathbb{R}^{n+k} : v_{n+j} - \sum_{i=1}^n \alpha_i^j(q) v_i = 0, 1 \leq j \leq k\}$. Θ is called the *connection form* of \mathcal{H} .

A horizontal curve γ in \mathbb{R}^{n+k} is an absolute continuous curve in \mathbb{R}^{n+k} whose tangent vector $\dot{\gamma}(t)$ belongs to $\mathcal{H}_{\gamma(t)}$ wherever it exists. Write $\gamma = (\gamma_1, \dots, \gamma_{n+k})$ in coordinates. Then γ is horizontal if and only if $\dot{\gamma}_{n+j} = \sum_{i=1}^n \alpha_i^j \dot{\gamma}_i$, $1 \leq j \leq k$, *a.e.* Fix a pair (m, q) where $m \in \mathbb{R}^n$, $q \in \mathbb{R}^{n+k}$, and $\pi(q) = m$. The *horizontal lift* (based at q) of a curve c in \mathbb{R}^n starting from m is defined as the unique horizontal curve γ in \mathbb{R}^{n+k} starting from q and satisfying $\pi(\gamma) = c$. If $c = (c_1, \dots, c_n)$ in coordinates, then $\gamma = (\gamma_1, \dots, \gamma_{n+k})$ is obtained by solving the following differential equations:

$$\gamma_1 = c_1, \dots, \gamma_n = c_n, \dot{\gamma}_{n+j} = \sum_{i=1}^n \alpha_i^j(\gamma) \dot{c}_i, \quad 1 \leq j \leq k. \quad (2.2)$$

If in particular c is a close loop, then γ starts and ends in the same fiber $\pi^{-1}(m)$, i.e., $\gamma(1) - \gamma(0)$ is of the form $(0, \dots, 0, h)$ for some $h \in \mathbb{R}^k$. We called h the *holonomy* of the loop c , which in general depends on the base point $q \in \pi^{-1}(m)$ of γ .

A sub-riemannian metric $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ on \mathcal{H} is a smooth assignment of inner products to the horizontal spaces \mathcal{H}_q . The length of a horizontal curve γ is measured as $L(\gamma) = \int_0^1 \|\dot{\gamma}\|_{\mathcal{H}} dt = \int_0^1 \langle \dot{\gamma}, \dot{\gamma} \rangle_{\mathcal{H}}^{1/2} dt$ under this metric. The sub-riemannian distance between two points in \mathbb{R}^{n+k} is the infimum of the length of all horizontal curves connecting them. Thus \mathcal{H} and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ specify a sub-riemannian geometry on \mathbb{R}^{n+k} .

2.2. Invariant distributions and sub-riemannian metrics. The distribution \mathcal{H} is called π -invariant, or simply invariant, with the bundle structure $\pi : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^n$ if its horizontal spaces are invariant along fibers. In terms of equation (2.1), this is equivalent to

$$\alpha_i^j(x_1, \dots, x_{n+k}) = \alpha_i^j(x_1, \dots, x_n), \quad 1 \leq i \leq n, 1 \leq j \leq k. \quad (2.3)$$

So we can think of α_i^j as functions on \mathbb{R}^n and define an \mathbb{R}^k -valued one-form on \mathbb{R}^n as

$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} \triangleq \begin{bmatrix} \sum_{i=1}^n \alpha_i^1 dx_i \\ \vdots \\ \sum_{i=1}^n \alpha_i^k dx_i \end{bmatrix}. \quad (2.4)$$

The holonomy of a loop c in \mathbb{R}^n is then independent of the starting point of its horizontal lift γ , and thus can be simply denoted by $h(c)$. Indeed, by (2.2) and an application of the Stokes' Theorem,

$$h(c) = \begin{bmatrix} \int_0^1 \dot{\gamma}_{n+1} dt \\ \vdots \\ \int_0^1 \dot{\gamma}_{n+k} dt \end{bmatrix} = \int_c \alpha = \int_S d\alpha = \int_S \beta, \quad (2.5)$$

where S is a two-dimensional submanifold immersed in \mathbb{R}^n whose boundary ∂S is exactly c , and β is the \mathbb{R}^k -valued two-form defined by

$$\beta \triangleq d\alpha = \sum_{1 \leq i, j \leq n} \beta_{ij} dx_i \wedge dx_j, \quad (2.6)$$

where β_{ij} , $1 \leq i, j \leq n$, are \mathbb{R}^k -valued functions on \mathbb{R}^n with $\beta_{ij} = -\beta_{ji}$. In (2.5), $h(c)$ is written as an integral of β over an arbitrary surface encircled by c .

For an invariant distribution \mathcal{H} , a sub-riemannian metric $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is called π -invariant (or invariant) with the bundle structure $\pi : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^n$ if it is also invariant along fibers. Invariant sub-riemannian metrics $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ on \mathcal{H} have a one-to-one correspondence with riemannian metrics $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ on the base space \mathbb{R}^n according to the following relation:

$$\langle h_q(u), h_q(v) \rangle_{\mathcal{H}} = \langle u, v \rangle_{\mathbb{R}^n}, \quad \forall u, v \in T_m \mathbb{R}^n, m \in \mathbb{R}^n, q \in \pi^{-1}(m). \quad (2.7)$$

Here $h_q : T_m \mathbb{R}^n \rightarrow \mathcal{H}_q$ is the horizontal lift operator defined as the inverse map of the linear isomorphism $d\pi_q : \mathcal{H}_q \rightarrow T_m \mathbb{R}^n$. We call $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ satisfying (2.7) the horizontal lift of $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$.

2.3. Asymptotic holonomy. Let $\mathcal{H} = \ker \Theta$ be a co-dimension k distribution on \mathbb{R}^{n+k} with the connection form Θ given in (2.1), and let $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ be a sub-riemannian metric on \mathcal{H} . In the rest of this paper, we shall assume that both \mathcal{H} and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ are invariant with the bundle structure $\pi : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^n$. Thus we can define the

forms α and β as in (2.4) and (2.6); and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the horizontal lift of a metric $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ on the base space \mathbb{R}^n . It should be pointed out, however, that the concepts of asymptotic rank and efficiency and some of their properties described below can be easily generalized to the non-invariant case.

Fix a point $m \in \mathbb{R}^n$ and a loop $c \neq 0$ in \mathbb{R}^n based at m . For each $\epsilon > 0$, denote by $c_\epsilon = m + \epsilon(c - m)$ the loop based at m obtained by scaling c by a factor of ϵ towards m , and let γ_ϵ be the horizontal lift of c_ϵ in \mathbb{R}^{n+k} based at $q \in \pi^{-1}(m) \subset \mathbb{R}^{n+k}$. We can define the following two quantities for c_ϵ : (1) its length $L(c_\epsilon) > 0$ is the length of c_ϵ in \mathbb{R}^n as measured by the metric $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$, which by (2.7) is also the length of the horizontal curve γ_ϵ as measured by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$; (2) its holonomy $h(c_\epsilon) \in \mathbb{R}^k$ is the vertical displacement between the two end points of γ_ϵ .

It is easy to see that $L(c_\epsilon)$ is of the same order of ϵ as $\epsilon \rightarrow 0$. In fact,

Lemma 1 $L(c_\epsilon) = a\epsilon + o(\epsilon)$, where $a \neq 0$ depends on $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ only through its restriction at m .

The continuity of $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ is needed to show the above claim. As for $h(c_\epsilon)$, we have

Lemma 2 $h(c_\epsilon) = \epsilon^{r(m)} \hat{h}(c) + o(\epsilon^{r(m)})$ for some constant $\hat{h}(c) \in \mathbb{R}^k$ and an integer

$$r(m) = \min\{l : \text{at least one } l\text{-th order partial derivative of } \beta \text{ at } m \text{ is nonzero}\} + 2. \quad (2.8)$$

Moreover, $\hat{h}(c) \neq 0$ for at least one loop c based at m .

Remark 3 The l -th order partial derivatives of β at m are terms of the form $\frac{\partial^l \beta(m)}{\partial x_1^{l_1} \cdots \partial x_n^{l_n}}$ for some integers l_1, \dots, l_n with $l_1 + \cdots + l_n = l$. Taking values in the set of \mathbb{R}^k -valued skew-symmetric 2-tensors on \mathbb{R}^n , each of these terms is zero if and only if all of its k components are zero.

Proof. Let S be a two dimensional submanifold of \mathbb{R}^n encircled by c . For each $\epsilon > 0$, denote by $S_\epsilon = m + \epsilon(S - m)$. Then $\partial S_\epsilon = c_\epsilon$ and $S_\epsilon \rightarrow m$ as $\epsilon \rightarrow 0$. Expanding $\beta = \sum_{1 \leq i, j \leq n} \beta_{ij} dx_i \wedge dx_j$ at $m = (m_1, \dots, m_n)$ in Taylor expansions and noticing the definition of $r(m)$ in (2.8), we have, for $x \in S_\epsilon$,

$$\begin{aligned} \beta(x) = & \sum_{l_1 + \cdots + l_n = r(m) - 2} \sum_{1 \leq i, j \leq n} \frac{\partial^{r(m) - 2} \beta_{ij}(m)}{\partial x_1^{l_1} \cdots \partial x_n^{l_n}} \frac{(x_1 - m_1)^{l_1} \cdots (x_n - m_n)^{l_n}}{l_1! \cdots l_n!} dx_i \wedge dx_j \\ & + o(\epsilon^{r(m) - 2}), \end{aligned}$$

So the dominating term of $h(c_\epsilon) = \int_{S_\epsilon} \beta$ as $\epsilon \rightarrow 0$ is

$$\sum_{l_1 + \cdots + l_n = r(m) - 2} \sum_{1 \leq i, j \leq n} \frac{1}{l_1! \cdots l_n!} \frac{\partial^{r(m) - 2} \beta_{ij}(m)}{\partial x_1^{l_1} \cdots \partial x_n^{l_n}} \int_{S_\epsilon} (x_1 - m_1)^{l_1} \cdots (x_n - m_n)^{l_n} dx_i \wedge dx_j,$$

which is exactly of the form $\epsilon^{r(m)} \hat{h}(c)$, where $\hat{h}(c)$ is given by

$$\sum_{l_1 + \cdots + l_n = r(m) - 2} \sum_{1 \leq i, j \leq n} \frac{1}{l_1! \cdots l_n!} \frac{\partial^{r(m) - 2} \beta_{ij}(m)}{\partial x_1^{l_1} \cdots \partial x_n^{l_n}} \int_S (x_1 - m_1)^{l_1} \cdots (x_n - m_n)^{l_n} dx_i \wedge dx_j. \quad (2.9)$$

It is easy to see that $\hat{h}(c) \neq 0$ for suitably chosen c and S . \square

Definition 1 $r(m)$ defined in (2.8) is called the rank of \mathcal{H} at $m \in M$, and

$$\eta(c) \triangleq \frac{\hat{h}(c)}{L^{r(m)}(c)} \in \mathbb{R}^k \quad (2.10)$$

is called the asymptotic holonomy of the loop c based at m , with $\hat{h}(c)$ defined in (2.9).

Despite its definition process, the rank $r(m)$ does not depend on the loop c . Indeed, by (2.8), $r(m)$ does not even depend on the subriemannian metric $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, and is solely determined by the distribution \mathcal{H} on the fibers over a neighborhood of m . Thus $r(m)$ is an intrinsic quantity of \mathcal{H} . On the other hand, the asymptotic holonomy $\eta(c)$ does depend on c ; and by Lemma 1 it is also affected by $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ (hence by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$) through its restriction at m . Let $A \in \mathbb{R}^{n \times n}$ be the positive definite matrix corresponding to the restriction of $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ at m , i.e.,

$$\langle u, v \rangle_{\mathbb{R}^n} = u^T A v, \quad (2.11)$$

for all $u, v \in T_m \mathbb{R}^n$. Then to compute $\eta(c)$ one can assume for convenience and without loss that $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ is given by A uniformly on \mathbb{R}^n , i.e., (2.11) holds for $u, v \in T_x \mathbb{R}^n$ for arbitrary $x \in \mathbb{R}^n$. Finally, the distribution \mathcal{H} affects $\eta(c)$ through $\hat{h}(c)$. By (2.9), in terms of computing $\eta(c)$, the form $\beta = \sum_{1 \leq i, j \leq n} \beta_{ij} dx_i \wedge dx_j$ can be replaced by the first nonvanishing term of its Taylor expansions: $\sum_{1 \leq i, j \leq n} \sum_{l_1 + \dots + l_n = r(m) - 2} \frac{1}{l_1! \dots l_n!} \frac{\partial^{r(m) - 2} \beta_{ij}(m)}{\partial x_1^{l_1} \dots \partial x_n^{l_n}} dx_i \wedge dx_j$, i.e., we can assume that the components of $\beta_{ij}(x)$ are homogeneous polynomials of degree $r(m) - 2$ in x with constant coefficients.

A direct consequence of Lemma 2 and Definition 1 is that, for any loop c based at m with $\eta(c) \neq 0$, $h(c_\epsilon) \sim \eta(c)[\epsilon L(c)]^{r(m)}$ as $\epsilon \rightarrow 0$. Here $a(\epsilon) \sim b(\epsilon)$ means that $\lim_{\epsilon \rightarrow 0} a(\epsilon)/b(\epsilon) = 1$ for functions a and b of $\epsilon > 0$ satisfying $b(\epsilon) \neq 0$ for $\epsilon \neq 0$.

Since $\hat{h}(\cdot)$ and $L^{r(m)}(\cdot)$ are both homogeneous of degree $r(m)$ in the scale ϵ of c_ϵ and are both invariant to reparameterizations of c , $\eta(c)$ has the following properties.

Lemma 3 (Invariance of asymptotic holonomy) *The asymptotic holonomy $\eta(c)$ of a loop c based at m is invariant to both scalings and reparameterizations of c , i.e.,*

- $\eta(c) = \eta(c_\epsilon)$ for any $\epsilon > 0$;
- $\eta(c \circ \rho) = \eta(c)$ for any orientation-preserving diffeomorphism $\rho : I \rightarrow I$.

As a result, $\eta(c)$ is a function of only the shape of the curve traversed by c , not of its size or the speed at which it is traversed. Indeed, $\eta(c)$ also remains unchanged if c is defined on a time interval $[0, T]$ other than $[0, 1]$. As argued in [8], these invariance properties are essential for a meaningful definition of the notion of holonomic efficiency. On the other hand, $\eta(c)$ changes sign if c is traversed in the reverse direction.

2.4. Optimization problem. In this section we define an optimization problem generalizing the one proposed in Section 1.

Definition 2 *A test function \mathcal{F} is a continuous map from \mathbb{R}^k to \mathbb{R} such that $\mathcal{F}(0) = 0$ and \mathcal{F} is linear along rays starting from the origin: $\mathcal{F}(\mu x) = \mu \mathcal{F}(x)$, $\forall x \in \mathbb{R}^k, \mu \geq 0$.*

The two important classes of \mathcal{F} considered in this paper are: (i) linear functions $\mathcal{F}(x) = \lambda^T x$ for some $\lambda \in \mathbb{R}^k$; and (ii) $\mathcal{F}(x) = \|x\|$ for some norm $\|\cdot\|$ on \mathbb{R}^k .

Problem 1 *Find the loop c in M based at $m \in M$ maximizing $\mathcal{F}[\eta(c)]$.*

Problem 1 originates as follows. Let $V : \mathbb{R}^k \rightarrow \mathbb{R}$ be a function with $V(0) = 0$ such that $V[h(c)]$ can be interpreted as the performance measure of the loop c . Then Problem 1 is the asymptotic version of the problem of finding the best performing loops c . Indeed, define the best performing c in the asymptotic sense as the ones for which $V[h(c_\epsilon)] \sim \mu[L(c_\epsilon)]^r$ as $\epsilon \rightarrow 0$ for the largest possible $\mu \in \mathbb{R}$ and the smallest possible integer r . Since $h(c_\epsilon) \sim \eta(c)[L(c_\epsilon)]^{r(m)}$ by the discussions in Section 2.3, we have $V[h(c_\epsilon)] \sim \mathcal{F}[\eta(c)][L(c_\epsilon)]^{r(m)}$ as $\epsilon \rightarrow 0$, where $\mathcal{F} : \mathbb{R}^k \rightarrow \mathbb{R}$ is defined as

$$\mathcal{F}(h) \triangleq \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} V(\epsilon h), \quad \forall h \in \mathbb{R}^k. \quad (2.12)$$

Hence, in the expression $V[h(c_\epsilon)] \sim \mu[L(c_\epsilon)]^r$, the smallest possible r is $r(m)$ and the largest possible μ is $\max\{\mathcal{F}[\eta(c)] : c\}$, both of which are achieved by solutions c to Problem 1, provided that the optimal $\mathcal{F}[\eta(c)] \neq 0$. Since normally V is differentiable along rays emitting from the origin, \mathcal{F} in (2.12) is well defined and satisfies the conditions in Definition 2. In particular, \mathcal{F} is linear if V is differentiable at 0, and $\mathcal{F} = V$ if V is a norm on \mathbb{R}^k .

In Problem 1 we assume that $\mathcal{F}[\eta(c)] > 0$ for at least one c to exclude the trivial solution $c \equiv 0$. This assumption is always satisfied for the examples in this paper.

Since $\mathcal{F}[\eta(c)] = \mathcal{F}[\hat{h}(c)]/L^{r(m)}(c)$, and both $\mathcal{F}[\hat{h}(\cdot)]$ and $L^{r(m)}(\cdot)$ are homogeneous of degree $r(m)$ in the scale ϵ of c_ϵ , solving Problem 1 is equivalent to solving any of the following variational problems.

$$\text{Find } c \text{ that minimize } L(c), \text{ subject to } \mathcal{F}[\hat{h}(c)] = 1; \quad (2.13)$$

$$\text{Find } c \text{ that maximize } \mathcal{F}[\hat{h}(c)], \text{ subject to } L(c) = 1; \quad (2.14)$$

$$\text{Find } c \text{ that minimize } E(c), \text{ subject to } \mathcal{F}[\hat{h}(c)] = 1; \quad (2.15)$$

$$\text{Find } c \text{ that maximize } \mathcal{F}[\hat{h}(c)] \text{ subject to } E(c) = 1. \quad (2.16)$$

Here $E(c) \triangleq \int_0^1 \|\dot{c}\|_{\mathbb{R}^n}^2 dt$ is the energy of c . Problems (2.14) and (2.16) are dual to problems (2.13) and (2.15), respectively. That problems (2.13) and (2.15), hence problems (2.14) and (2.16), are equivalent is because of the inequality $E(c) \geq L^2(c)$ with equality if and only if c has constant speed; thus solutions to the latter are necessarily solutions to the former parameterized with constant speed. In this paper, to avoid the ambiguity of parameterizations, we will study problems (2.15) and (2.16).

By the discussion immediately after Definition 1, to solve these problems we can assume the following without loss:

Assumption 1 *Assume that*

1. $m = 0$ is the origin of \mathbb{R}^n ;
2. $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ is given by a positive definite $A \in \mathbb{R}^{n \times n}$ on \mathbb{R}^n . In fact, after a change of orthonormal coordinates, we can assume that $A = I_n$, i.e., $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ is the standard metric on \mathbb{R}^n ;
3. $\beta = \sum_{1 \leq i, j \leq n} \beta_{ij} dx_i \wedge dx_j \neq 0$, where $\beta_{ij} = -\beta_{ji}$, and the components of $\beta_{ij}(x) \in \mathbb{R}^k$ are homogeneous polynomials of degree $r(0) - 2$ in $x \in \mathbb{R}^n$ with constant coefficients.

Thus $L(c)$ and $E(c)$ are the standard arc length and energy of the loop c based at 0; and $\hat{h}(c)$ in (2.9) coincides with $h(c)$. Problem (2.15) and (2.16) can then be reformulated respectively as

Problem 2 (Generalized Isoholonomic Problem) Find c with $\mathcal{F}[h(c)] = 1$ minimizing $E(c)$.

Problem 3 (Generalized Isoperimetric Problem) Find c with $E(c) = 1$ maximizing $\mathcal{F}[h(c)]$.

For an even more general formulation of the isoholonomic problem, see [4].

Solutions to the above two problems are the same up to a scaling. To derive their equations, note that the solutions to Problem 2 also solve the following problem for some proper $h_0 \in \mathbb{R}^k$:

$$\text{Find } c \text{ with a fixed } h(c) = h_0 \text{ that minimize } E(c). \quad (2.17)$$

It is shown in [10] that the solutions to problem (2.17) satisfy

$$\ddot{c} = -i_{\dot{c}}(\lambda^T \beta), \quad (2.18)$$

for some constant $\lambda \in \mathbb{R}^k$. Here $\lambda^T \beta$ is an \mathbb{R} -valued two-form, and $i_{\dot{c}}(\lambda^T \beta) \triangleq \lambda^T \beta(\dot{c}, \cdot)$ is a one-form on \mathbb{R}^n that we identify as a vector in \mathbb{R}^n via the canonical metric. Write $\beta = \sum_{1 \leq i, j \leq n} \beta_{ij} dx_i \wedge dx_j$ in coordinates. Then equation (2.18) is equivalent to

$$\ddot{c} = Z \dot{c}, \quad (2.19)$$

where Z is the skew-symmetric matrix

$$Z = \begin{bmatrix} 2\lambda^T \beta_{11} & \cdots & 2\lambda^T \beta_{1n} \\ \vdots & \ddots & \vdots \\ 2\lambda^T \beta_{n1} & \cdots & 2\lambda^T \beta_{nn} \end{bmatrix} \quad (2.20)$$

whose components are homogeneous polynomials of degree $r(0) - 2$ in x with constant coefficients. Indeed, (2.19) describes the motions of a particle of unit mass and unit charge moving in a magnetic field given by Z on \mathbb{R}^n when $n = 2, 3$ (see [18, 4]). Equation (2.19) can also be derived from the Pontryagin Maximum Principle [19] by formulating problem (2.17) as an optimal control problem.

Equation (2.19) does not solve Problem 2 completely, as λ is unknown and we are interested only in those solutions that start and end in the origin. Thus we still need to determine λ and the appropriate initial condition $\dot{c}(0)$ such that $c(0) = c(1) = 0$, which is often a non-trivial task.

3. Rank Two Case. We first study the solution of Problem 2 in the simplest case, namely, the rank two case. In this case, when the test function is linear, it is a well known fact that the optimal loops are simple circles, and straightforward procedures exist to find these optimal circles (Section 3.1). Indeed, when $n = 2$, the rank-two isoholonomic problem degenerates into the classical isoperimetric problem. See the discussion in [10, Chap. 1]. Our contribution in this section lies in the generalization of this result to the case of convex test functions (Theorem 1). In addition, for a class of convex but not linear test functions, in Section 3.2 we propose iterative procedures to solve for the optimal circles, which is a much more difficult problem than in the linear test function case.

Suppose that $r(0) = 2$. Then $\beta = \sum_{1 \leq i, j \leq n} \beta_{ij} dx_i \wedge dx_j \neq 0$ for constants $\beta_{ij} = -\beta_{ji}$. Z defined in (2.20) is a constant matrix. Being skew-symmetric, Z admits a decomposition of the form

$$Z = Q \cdot \text{diag} \left(\begin{bmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 & -\sigma_l \\ \sigma_l & 0 \end{bmatrix}, 0, \dots, 0 \right) \cdot Q^T,$$

where $Q \in \mathbb{R}^{n \times n}$ is orthonormal, $\sigma_1 \geq \dots \geq \sigma_l > 0$ for some integer l with $2l = \text{rank}(Z)$. After an orthonormal coordinate transformation $y = Q^T x$, (2.19) becomes

$$\begin{cases} \begin{bmatrix} \ddot{y}_{2p-1} \\ \ddot{y}_{2p} \end{bmatrix} = \begin{bmatrix} 0 & -\sigma_p \\ \sigma_p & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_{2p-1} \\ \dot{y}_{2p} \end{bmatrix}, & p = 1, \dots, l, \\ \ddot{y}_p = 0, & p = 2l + 1, \dots, n. \end{cases}$$

Solutions to this equation that start and end at the origin are necessarily of the form

$$c(t) = [a_1(1 - \cos(2n_1\pi t)), -a_1 \sin(2n_1\pi t), \dots, a_l(1 - \cos(2n_l\pi t)), -a_l \sin(2n_l\pi t), 0, \dots, 0]^T \quad (3.1)$$

for some $a_1, \dots, a_l \in \mathbb{R}$, and some $n_1, \dots, n_l \in \mathbb{N}$ with $\sigma_1 = 2n_1\pi, \dots, \sigma_l = 2n_l\pi$. Note that we can assume without loss that $n_p, p = 1, \dots, l$, are all distinct. Otherwise, for example, if $n_1 = n_2$, then a suitable change of orthonormal coordinates within the 4-subspace spanned by the y_1, \dots, y_4 axes can transform c into the form

$$[a(1 - \cos(2n_1\pi t)), -a \sin(2n_1\pi t), 0, 0, a_3(1 - \cos(2n_3\pi t)), -a_3 \sin(2n_3\pi t), \dots]^T$$

with $a = \sqrt{a_1^2 + a_2^2}$. This step can be repeated until all n_p are distinct eventually, resulting in a curve of the form

$$c \text{ is given in (3.1) for some } 1 \leq l \leq \lceil n/2 \rceil \text{ and distinct } n_1, \dots, n_l \neq 0. \quad (3.2)$$

Curves of the form (3.2) in some orthonormal coordinates of \mathbb{R}^n are called *mixed circles*. If in particular $l = 1$ and $n_1 = 1$ in (3.2), the resulting curves are called *simple circles*, which are planar circle in \mathbb{R}^n traversed exactly once.

It is seen from the above that the solutions to Problem 2 are mixed circles. In the case of convex \mathcal{F} , the solution can be further simplified.

Theorem 1 *Suppose that $\mathcal{F} : \mathbb{R}^k \rightarrow \mathbb{R}$ is convex. Then there is at least one simple circle solution to Problem 2 (Problem 3).*

The result of Theorem 1 has been well known in the literature when \mathcal{F} is a linear function. To prove it in the general convex \mathcal{F} case, we first introduce an intermediate result, which is reformulated from the arguments in [10, Sec. 12.3.5]. Consider a mixed circle c of the form (3.2) in some orthonormal coordinates (y_1, \dots, y_n) . For each $p = 1, \dots, l$, denote by $c_{(p)}$ the orthogonal projection of c onto the plane spanned by the y_{2p-1} and y_{2p} axes, which is a planar circle traversed n_p times.

Lemma 4 $h(c) = h(c_{(1)}) + \dots + h(c_{(l)})$.

Proof. Write $\beta = \sum_{1 \leq i, j \leq n} \hat{\beta}_{ij} dy_i \wedge dy_j$ in the new coordinates, with constants $\hat{\beta}_{ij} = -\hat{\beta}_{ji}$. Define $\alpha \triangleq \sum_{i, j=1}^n \hat{\beta}_{ij} y_i dy_j$. Then $d\alpha = \beta$, and, by (2.5), $h(c) = \int_c \alpha = \sum_{i, j=1}^n \hat{\beta}_{ij} \int_c y_i dy_j$. Note that because of the special form of c in (3.1) and (3.2), unless $\{i, j\} = \{2p-1, 2p\}$ for some $p = 1, \dots, l$, we must have $\int_c y_i dy_j = 0$, since the integral of the product of two periodic sine or cosine functions with different frequencies is zero. As a result,

$$h(c) = \sum_{p=1}^l \int_c \hat{\beta}_{2p-1, 2p} (y_{2p-1} dy_{2p} - y_{2p} dy_{2p-1}) = \sum_{p=1}^l \int_{c_{(p)}} \alpha = \sum_{p=1}^l h(c_{(p)}),$$

which proves the desired conclusion. \square

Now define three subsets of \mathbb{R}^k :

$$\begin{aligned}\mathcal{B}_0 &= \{h(c) : c \text{ is a loop with } E(c) \leq 1\}, \\ \mathcal{B}_1 &= \{h(c) : c \text{ is a mixed circle with } E(c) \leq 1\}, \\ \mathcal{B}_2 &= \{h(c) : c \text{ is a simple circle with } E(c) \leq 1\}.\end{aligned}$$

\mathcal{B}_0 is the set of holonomy achievable by loops with energy no larger than 1, and is the intersection of the unit sub-riemannian ball centered at 0 with the fiber \mathbb{R}^k through 0. Obviously, \mathcal{B}_0 is star-shaped ($h \in \mathcal{B}_0$ implies $\mu h \in \mathcal{B}_0$ for $\mu \in [0, 1]$) and symmetric ($h \in \mathcal{B}_0$ implies $-h \in \mathcal{B}_0$).

Since our previous analysis shows that every holonomy achievable by a loop c can be achieved by a mixed circle with no more energy, we have $\mathcal{B}_0 = \mathcal{B}_1$. Obviously, $\mathcal{B}_2 \subset \mathcal{B}_1$. But Lemma 4 implies

Lemma 5 *Co(\mathcal{B}_1) = Co(\mathcal{B}_2), i.e., \mathcal{B}_1 and \mathcal{B}_2 span the same convex hull.*

Proof. Since \mathcal{B}_1 and \mathcal{B}_2 are closed sets, it suffices to show that for any $\lambda \in \mathbb{R}^k$, $\sup(\lambda^T \mathcal{B}_1) = \sup(\lambda^T \mathcal{B}_2)$. Suppose that $\sup(\lambda^T \mathcal{B}_1)$ is achieved at $h(c) \in \mathcal{B}_1$ for a mixed circle c of the form (3.2) in some orthonormal coordinates with $a_1, \dots, a_l \neq 0$ and $E(c) \leq 1$. Note that $\lambda^T h(c) \geq 0$ since $0 \in \mathcal{B}_1$. By Lemma 4, $h(c) = h(c_{(1)}) + \dots + h(c_{(l)})$, and

$$\frac{\lambda^T h(c)}{E(c)} = \frac{\lambda^T h(c_{(1)}) + \dots + \lambda^T h(c_{(l)})}{E(c_{(1)}) + \dots + E(c_{(l)})} \leq \max_{1 \leq p \leq l} \frac{\lambda^T h(c_{(p)})}{E(c_{(p)})}.$$

Suppose that the maximum in the above equation is achieved at p . Then the simple circle $\hat{c}_{(p)}(t) = \frac{1}{2\pi}(0, \dots, 0, 1 - \cos(2\pi t), \sin(2\pi t), 0, \dots, 0)$ traversing the (scaled) image of $c_{(p)}$ exactly once has unit energy and holonomy $h(\hat{c}_{(p)}) = n_p h(c_{(p)})/E(c_{(p)})$. From the above inequality, we have

$$\lambda^T h(\hat{c}_{(p)}) = n_p \lambda^T h(c_{(p)})/E(c_{(p)}) \geq \lambda^T h(c_{(p)})/E(c_{(p)}) \geq \lambda^T h(c)/E(c) \geq \lambda^T h(c) \geq 0.$$

Since $h(\hat{c}_{(p)}) \in \mathcal{B}_2 \subset \mathcal{B}_1$, $\sup(\lambda^T \mathcal{B}_2) = \sup(\lambda^T \mathcal{B}_1)$. Therefore, $\text{Co}(\mathcal{B}_1) = \text{Co}(\mathcal{B}_2)$. \square

Example 1 (Brockett [13]) Consider the total space $\mathbb{R}^n \oplus \mathfrak{so}_n \simeq \mathbb{R}^{n(n+1)/2}$, whose elements are (x, A) with $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ skew symmetric matrices. Let \mathcal{H} be the co-dimension $\frac{n(n-1)}{2}$ distribution invariant with $\pi : \mathbb{R}^n \oplus \mathfrak{so}_n \rightarrow \mathbb{R}^n$ given by the \mathfrak{so}_n -valued one-form $\alpha = (x \cdot dx^T - dx \cdot x^T)/2$. Thus $\beta = dx \wedge dx^T$. It is easy to verify that \mathcal{B}_1 consists of all matrices of the form

$$Q \cdot \text{diag} \left(\begin{bmatrix} 0 & n_1 \pi a_1^2 \\ -n_1 \pi a_1^2 & 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 & n_l \pi a_l^2 \\ -n_l \pi a_l^2 & 0 \end{bmatrix}, 0, \dots, 0 \right) \cdot Q^T$$

for some $Q \in \mathbb{O}_n$, $1 \leq l \leq \lfloor \frac{n}{2} \rfloor$, $a_1, \dots, a_l \in \mathbb{R}$, and some distinct $n_1, \dots, n_l \in \mathbb{N}$ such that $4n_1 \pi^2 a_1^2 + \dots + 4n_l \pi^2 a_l^2 \leq 1$. On the other hand, \mathcal{B}_2 is

$$\mathcal{B}_2 = \left\{ Q \cdot \text{diag} \left(\begin{bmatrix} 0 & \pi a_1^2 \\ -\pi a_1^2 & 0 \end{bmatrix}, 0, \dots, 0 \right) \cdot Q^T : Q \in \mathbb{O}_n, 4\pi^2 a_1^2 \leq 1 \right\}.$$

Note that $\mathcal{B}_1 \neq \mathcal{B}_2$ since matrices in \mathcal{B}_2 have rank at most 2 while the rank of matrices in \mathcal{B}_1 can be any even number between 0 and n . In other words, certain holonomy in \mathfrak{so}_n can be achieved by mixed circles but not simple circles.

This example is universal for the problem studied in this section: Any other distribution invariant with $\pi : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^n$ specified by a form β with non-trivial constant coefficients is induced from \mathcal{H} in this example by a linear transformation $\mathbb{R}^n \oplus \mathfrak{so}_n \rightarrow \mathbb{R}^{n+k}$ that leaves \mathbb{R}^n invariant and transforms \mathfrak{so}_n to \mathbb{R}^k properly. Therefore, as an alternative it suffices to prove Lemma 5 for \mathcal{B}_1 and \mathcal{B}_2 in this particular example only, since convexity is preserved by linear transformations.

Theorem 1 then follows easily from Lemma 5. In fact, Problem 3 is equivalent to finding $\max\{\mathcal{F}(h) : h \in \mathcal{B}_0\} = \max\{\mathcal{F}(h) : h \in \mathcal{B}_1\}$. By Lemma 5 and the convexity of \mathcal{F} , $\max\{\mathcal{F}(h) : h \in \mathcal{B}_1\} = \max\{\mathcal{F}(h) : h \in \mathcal{B}_2\}$. So there is at least a simple circle solution to Problem 3. Since solutions to Problem 2 are scaled versions of solutions to Problem 3, this proves Theorem 1.

Remark 4 *For the above reasoning to hold, we only need \mathcal{F} to be quasi-convex instead of convex (a function $f : S \rightarrow \mathbb{R}$ defined a convex subset of \mathbb{R}^n is called quasi-convex if each of its sub-level set of the form $\{x : f(x) < a\}$ is convex for $a \in \mathbb{R}$). However, these two properties are equivalent due to the linearity of \mathcal{F} along rays.*

Now consider Problem 3. By Theorem 1, there is a solution of the form

$$c(t) = \frac{1}{2\pi}[(1 - \cos(2\pi t))u + \sin(2\pi t)v] \quad (3.3)$$

for a pair of orthonormal vectors u and v in \mathbb{R}^n . For c given by (3.3), direct computation shows that $h(c) = \frac{1}{4\pi}\beta(u, v)$; thus

$$\mathcal{F}[h(c)] = \frac{\mathcal{F}[\beta(u, v)]}{4\pi}. \quad (3.4)$$

So to solve Problem 3 it suffices to find the pair (u, v) that maximizes $\mathcal{F}[\beta(u, v)]$.

In the following, we will outline the procedures to determine the simple circle solutions in the cases when \mathcal{F} is linear and when \mathcal{F} is a norm.

3.1. Solution circles when \mathcal{F} is linear. In this section, we briefly outline the procedure for finding the solution circles when \mathcal{F} is a linear function. Such a procedure is well known in the literature; and we include it here for completeness. Suppose that $\mathcal{F}(h) = \rho^x h$ is linear for some constant $\rho \in \mathbb{R}^k$. Then

$$\mathcal{F}[\beta(u, v)] = \rho^x \beta(u, v) = u^x Z_0 v,$$

where Z_0 is the skew symmetric matrix defined by

$$Z_0 = \begin{bmatrix} 2\rho^x \beta_{11} & \cdots & 2\rho^x \beta_{1n} \\ \vdots & \ddots & \vdots \\ 2\rho^x \beta_{n1} & \cdots & 2\rho^x \beta_{nn} \end{bmatrix}.$$

Denote by $\sigma_1(Z_0)$ the largest singular value of Z_0 . Then it is a well known fact in linear algebra that the orthonormal u and v that maximize $u^x Z_0 v$ must be the left and right singular vectors of Z_0 corresponding to the singular value $\sigma_1(Z_0)$, i.e.,

$$Z_0 u = \sigma_1(Z_0)v, \quad Z_0 v = -\sigma_1(Z_0)u. \quad (3.5)$$

Together, u and v span a two dimensional subspace of \mathbb{R}^n invariant under Z_0 . A solution to Problem 3 is then given by (3.3). A solution to Problem 2 is a scaled

version of (3.3). These solutions are well known for the micro-organism swimming problem, for example, in [8] where $\mathbb{R}^k = \mathbb{R}^3$ is the space of translations of the micro-organism and ρ is aligned with the positive z -axis.

Note that (u, v) satisfying (3.5) is in general not unique for two reasons: Z_0 could have multiple singular values equal to $\sigma_1(Z_0)$; and even if this is not the case, a rotation of u and v within the plane they span will yield a new orthonormal pair satisfying (3.5). As a result, any simple circle of unit energy through the origin and contained in an invariant plane of Z_0 corresponding to $\sigma_1(Z_0)$ will solve Problem 3.

3.2. Solution circles when \mathcal{F} is a norm. Suppose that $\mathcal{F} = \|\cdot\|$ is a norm on \mathbb{R}^k . Finding the optimal circle is considerably more difficult in this case. In this section, we will propose a novel procedure, Algorithm 1, to determine the orthonormal pair (u, v) that maximizes $\mathcal{F}[\beta(u, v)] = \|\beta(u, v)\|$. A solution to Problem 3 is then given by a simple circle contained in the plane spanned by u and v .

First of all, for each $p = 1, \dots, k$, define Z^p as the skew symmetric matrix

$$Z^p \triangleq \begin{bmatrix} 2\beta_{11}^p & \cdots & 2\beta_{1n}^p \\ \vdots & \ddots & \vdots \\ 2\beta_{n1}^p & \cdots & 2\beta_{nn}^p \end{bmatrix},$$

where β_{ij}^p is the p -th component of $\beta_{ij} \in \mathbb{R}^k$, $1 \leq i, j \leq n$. Let $\sigma_1(Z^p)$ be the largest singular value of Z^p , and let (u_p, v_p) be a pair of left and right singular vectors of Z^p corresponding to $\sigma_1(Z^p)$.

The solution is simple when $\|\cdot\|$ is the L^1 - or L^∞ -norm. So we will simply point out the results. If \mathcal{F} is the L^∞ -norm, the pair (u, v) that maximizes $\|\beta(u, v)\|$ is the pair (u_p, v_p) for a p with the largest $\sigma_1(Z^p)$. If \mathcal{F} is the L^1 -norm, then define $\mathcal{A} \triangleq \{\pm Z^1 \pm Z^2 \pm \dots \pm Z^k\}$, and choose a $Z \in \mathcal{A}$ with the largest $\sigma_1(Z)$. The pair (u, v) that maximizes $\|\beta(u, v)\|$ is then given by a pair of left and right singular vectors of Z corresponding to $\sigma_1(Z)$.

We focus on the more interesting case where $\|\cdot\|$ is the L^2 -norm in the rest of this section. In this case, since β is anti-symmetric, the pair (u, v) maximizing $\|\beta(u, v)\|$ subject to $\|u\| = \|v\| = 1$ will automatically be orthogonal to each other. So we might as well drop the orthogonality constraint. Write

$$\|\beta(u, v)\|^2 = \sum_{p=1}^k (u^T Z^p v)^2 = u^T \left(\sum_{p=1}^k Z^p v v^T (Z^p)^T \right) u = v^T \left(\sum_{p=1}^k (Z^p)^T u u^T Z^p \right) v.$$

Therefore, to maximize $\|\beta(u, v)\|$ under the constraint that $\|u\| = \|v\| = 1$, we need

- (i) u is an eigenvector of $\sum_{p=1}^k Z^p v v^T (Z^p)^T$ for its largest eigenvalue;
- (ii) v is an eigenvector of $\sum_{p=1}^k (Z^p)^T u u^T Z^p$ for its largest eigenvalue.

These two conditions hint at the following iterative algorithm.

Algorithm 1 Choose some initial u and v in \mathbb{R}^n such that $\|u\| = \|v\| = 1$.

1. Let u be a unit eigenvector of $\sum_{p=1}^k Z^p v v^T (Z^p)^T$ for its largest eigenvalue;
2. Let v be a unit eigenvector of $\sum_{p=1}^k (Z^p)^T u u^T Z^p$ for its largest eigenvalue;
3. Repeat steps 1 and 2 until some convergence criteria is satisfied, for example, when the changes in u, v in consecutive steps are below given threshold.

The value of $\|\beta(u, v)\|$ increases with each iteration, and, barring the occurrence of cycles, u and v will converge to a pair satisfying conditions (i) and (ii). However, as

these are only necessary conditions, the convergence property to the global solutions is still an issue to be resolved.

Remark 5 *Bounds on $\max\{\|\beta(u, v)\| : \|u\| = \|v\| = 1\}$ can be obtained as follows. For each $p = 1, \dots, k$, let the column vectors of Z^p from left to right be stacked from top to bottom into a single column vector $\mathbf{z}_p \in \mathbb{R}^{n^2}$. In addition, for $u = (u_1, \dots, u_n)$ and $v = (v_1, \dots, v_n)$ in \mathbb{R}^n , denote by $u \otimes v$ the vector $(u_1 v_1, \dots, u_1 v_n, \dots, u_n v_1, \dots, u_n v_n) \in \mathbb{R}^{n^2}$. Then $\mathbf{z}_p^T(u \otimes v) = u^T Z^p v$, and*

$$\beta(u, v) = \begin{bmatrix} u^T Z^1 v \\ \vdots \\ u^T Z^k v \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1^T(u \otimes v) \\ \vdots \\ \mathbf{z}_k^T(u \otimes v) \end{bmatrix} = \mathbf{Z}(u \otimes v),$$

where $\mathbf{Z} \in \mathbb{R}^{k \times n^2}$ is defined by $\mathbf{Z} = [\mathbf{z}_1 \ \cdots \ \mathbf{z}_k]^T$. Since $u \otimes v$ is a unit vector in \mathbb{R}^{n^2} for unit u, v , we have

$$\|\beta(u, v)\| = \|\mathbf{Z}(u \otimes v)\| \leq \sigma_1(\mathbf{Z}). \quad (3.6)$$

In general, $\{u \otimes v : \|u\| = \|v\| = 1\}$ is a proper subset of the unit sphere in \mathbb{R}^{n^2} . So the bound (3.6) is not strict. By re-arranging Z^p in different ways, we can obtain other bounds similar to (3.6). See [20] for more on the singular value decomposition of multilinear tensors.

In the case when the base space has dimension three, i.e., when $n = 3$, the solution is especially simple. In fact, for each $p = 1, \dots, k$, since $Z^p \in \mathbb{R}^{3 \times 3}$ is skew-symmetric, we can find $z_p \in \mathbb{R}^3$ such that $Z^p v = v \times z_p, \forall v \in \mathbb{R}^3$. Here \times denotes the cross product of vectors in \mathbb{R}^3 . Therefore,

$$\beta(u, v) = \begin{bmatrix} u^T Z^1 v \\ \vdots \\ u^T Z^k v \end{bmatrix} = \begin{bmatrix} u^T(v \times z_1) \\ \vdots \\ u^T(v \times z_k) \end{bmatrix} = \begin{bmatrix} z_1^T(u \times v) \\ \vdots \\ z_k^T(u \times v) \end{bmatrix} = \mathbf{Z}(u \times v), \quad (3.7)$$

where $\mathbf{Z} \in \mathbb{R}^{k \times 3}$ is defined by $\mathbf{Z} \triangleq [z_1 \ \cdots \ z_k]^T$. Note that the set of $u \times v$ for unit u and v is exactly the unit ball in \mathbb{R}^3 . Therefore,

$$\|\beta(u, v)\| = \|\mathbf{Z}(u \times v)\| \leq \sigma_1(\mathbf{Z}),$$

with exact equality achieved by any orthonormal pair u and v with $u \times v = w$, where $w \in \mathbb{R}^3$ is a unit right singular vector of \mathbf{Z} corresponding to $\sigma_1(\mathbf{Z})$.

4. Higher Rank Case. Solving Problems 2 and 3 in the higher rank case is much more difficult than in the rank two case, as analytic characterization of solutions is in general not available. In this section, however, we shall study a special class of higher rank problems for which analytical characterization is possible. The result will then be applied in Section 5 to the linked-mass system in Section 1.1.

Consider the following co-dimensional one distribution \mathcal{H} on \mathbb{R}^3 with base space \mathbb{R}^2 . The forms α and β specifying \mathcal{H} as in (2.4) and (2.6) are respectively

$$\alpha = x_1^r dx_2, \quad \beta = r x_1^{r-1} dx_1 \wedge dx_2, \quad (4.1)$$

for some integer $r \geq 2$. When $r = 2$, this distribution is called the Martinet distribution and has been well studied in the sub-riemmanian geometry and optimal control

literature (e.g. [21, 22]). By Lemma 2, the rank of \mathcal{H} at the origin is $r + 1$. Suppose that the metric $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is obtained by lifting the the standard metric on \mathbb{R}^2 , and that $\mathcal{F} : \mathbb{R} \rightarrow \mathbb{R}$ is the identity map. So Assumption 1 in Section 2.4 is satisfied.

For a loop c in \mathbb{R}^2 based at 0 enclosing a surface S , $h(c) = \int_c x_1^r dx_2 = \int_S r x_1^{r-1} dx_1 \wedge dx_2$. Problem 2 then reduces to:

$$\text{Find } c \text{ with } \int_c x_1^r dx_2 = 1 \text{ that minimize } E(c). \quad (4.2)$$

Lemma 6 *There is a solution c to problem (4.2) such that*

1. c is contained exclusively in the closed right half plane;
2. c has no self-crossing, and encloses a convex region S ;
3. S is symmetric with respect to the x_1 -axis.

To prove each claim one shows that for a loop c not satisfying the condition, a better loop can be obtained by transforming c properly. As an example, for claim 2 one can “flip” outside a certain segment of a non-convex c contained strictly in its convex hull to obtain a loop with the same energy but a larger holonomy. We omit the proof here.

Let c be a solution to problem (4.2) satisfying the conditions in Lemma 6. Because of the symmetry, it suffices to study the first half of c only, which starts from the origin at time 0, follows the graph of a convex function f below the x_1 -axis during $[0, 1/2]$, and reaches a point $(a, 0)$ with $a > 0$ on the x_1 -axis at time $1/2$. Such a $c = (x_1, x_2)$ must solve the following optimal control problem:

$$\begin{aligned} & \text{Minimize} && \frac{1}{2} \int_0^{\frac{1}{2}} (u_1^2 + u_2^2) dt && (4.3) \\ & \text{subject to} && \begin{cases} \dot{x}_1 = u_1, & x_1(0) = 0, & x_1(\frac{1}{2}) = a, \\ \dot{x}_2 = u_2, & x_2(0) = 0, & x_2(\frac{1}{2}) = 0, \\ \dot{x}_3 = x_1^r u_2, & x_3(0) = 0, & x_3(\frac{1}{2}) = \frac{1}{2}. \end{cases} \end{aligned}$$

Note that a is a parameter to be determined later so that $\dot{x}_1(\frac{1}{2}) = 0$; thus the two halves of c can be pieced together smoothly. The boundary condition $x_3(\frac{1}{2}) = \frac{1}{2}$ is imposed to ensure that $\int_c x_1^r dx_2 = 1$.

Define the Hamiltonian:

$$H(\lambda_1, \lambda_2, \lambda_3, x_1, x_2, x_3) = \frac{u_1^2 + u_2^2}{2} + \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 x_1^r u_2.$$

By the maximum principle [19], u_1, u_2 for the optimal solution can be determined as

$$u_1 = \operatorname{argmin}_{u_1} H = -\lambda_1, \quad (4.4)$$

$$u_2 = \operatorname{argmin}_{u_2} H = -\lambda_2 - \lambda_3 x_1^r, \quad (4.5)$$

while λ_i , $i = 1, 2, 3$, satisfy

$$\dot{\lambda}_1 = -\frac{\partial H}{\partial x_1} = -r\lambda_3 x_1^{r-1} u_2, \quad \dot{\lambda}_2 = -\frac{\partial H}{\partial x_2} = 0, \quad \dot{\lambda}_3 = -\frac{\partial H}{\partial x_3} = 0.$$

Thus λ_2 and λ_3 are constant. Their signs can be determined as $\lambda_2 \geq 0$ and $\lambda_3 < 0$. In fact, at time $t = 0$ in (4.5), we have $x_1 = 0$, hence $-\lambda_2 = u_2(0) = \dot{x}_2(0) \leq 0$ by our assumption that the curve f is below the x_1 -axis during $[0, 1/2]$, i.e., $\lambda_2 \geq$

0. Denote $\tau \in (0, \frac{1}{2})$ the time when x_2 achieves its minimum during $[0, \frac{1}{2}]$. Thus, $\dot{x}_2(\tau) = -\lambda_2 - \lambda_3 x_1^r(\tau) = 0$, which is possible only if $\lambda_3 \leq 0$. Moreover, $\lambda_3 \neq 0$, for otherwise $\dot{x}_2 = u_2 = -\lambda_2$ is constant zero, an obvious contradiction.

Note that $\ddot{x}_1 = \dot{u}_1 = -\dot{\lambda}_1 = r\lambda_3 x_1^{r-1} u_2 = -r\lambda_3 x_1^{r-1} (\lambda_2 + \lambda_3 x_1^r)$. Hence,

$$d(\dot{x}_1^2) = 2\ddot{x}_1 dx_1 = -2r\lambda_3 x_1^{r-1} (\lambda_2 + \lambda_3 x_1^r) \dot{x}_1 = -d(\lambda_2 + \lambda_3 x_1^r)^2.$$

The integrability of the above equation is a key result that can greatly reduce the complexity of solving the optimal control problem (4.3). After integration, we obtain

$$\dot{x}_1^2 = \lambda_3^2 C^2 - (\lambda_2 + \lambda_3 x_1^r)^2$$

for some constant $C > 0$. Therefore,

$$\dot{x}_1 = -\lambda_3 \sqrt{C^2 - (\lambda_2/\lambda_3 + x_1^r)^2}. \quad (4.6)$$

Note that $|C| \geq \lambda_2/\lambda_3$, for otherwise \dot{x}_1 is not defined at time 0. Since $x_1 = a$ and $\dot{x}_1 = 0$ at time $t = \frac{1}{2}$, a can be determined as

$$a = (C - \lambda_2/\lambda_3)^{1/r}. \quad (4.7)$$

The graph of the function f that c follows during $[0, \frac{1}{2}]$ can be derived directly. Dividing $\dot{x}_2 = u_2 = -(\lambda_2 + \lambda_3 x_1^r)$ by (4.6), we have

$$\frac{dx_2}{dx_1} = \frac{\lambda_2/\lambda_3 + x_1^r}{\sqrt{C^2 - (\lambda_2/\lambda_3 + x_1^r)^2}}, \quad 0 \leq x_1 \leq a. \quad (4.8)$$

Integrating the above equation with respect to x_1 will yield x_2 as a function of $x_1 \in [0, \frac{1}{2}]$, namely, the graph of the function f . It remains to determine the unknown parameters λ_2/λ_3 and C in (4.7). The boundary conditions $x_2(\frac{1}{2}) = 0$ and $x_3(\frac{1}{2}) = \frac{1}{2}$ imply respectively that

$$\int_0^a \frac{\lambda_2/\lambda_3 + x_1^r}{\sqrt{C^2 - (\lambda_2/\lambda_3 + x_1^r)^2}} dx_1 = 0, \quad (4.9)$$

$$\int_0^a \frac{x_1^r (\lambda_2/\lambda_3 + x_1^r)}{\sqrt{C^2 - (\lambda_2/\lambda_3 + x_1^r)^2}} dx_1 = \frac{1}{2}. \quad (4.10)$$

The procedures to determine λ_2/λ_3 and C satisfying the above conditions are listed below:

1. Choose any fixed λ_2/λ_3 , say, $\lambda_2/\lambda_3 = \kappa_0 < 0$;
2. Find C so that (4.9) is satisfied, say, $C = C_0$. Note that a in (4.9) is determined by (4.7);
3. Use κ_0 and C_0 in (4.7) to compute an a , say, $a = a_0$;
4. Use κ_0 , C_0 , and a_0 in (4.8) to integrate for a function $x_2 = g(x_1)$ on $[0, a_0]$;

The function g obtained so far is in general not the desired function f , since constraint (4.10) may not be satisfied. However, f can be obtained from g by a proper scaling. In fact, define

$$\mu = \left[2 \int_0^{a_0} \frac{x_1^r (\kappa_0 + x_1^r)}{\sqrt{C_0^2 - (\kappa_0 + x_1^r)^2}} dx_1 \right]^{1/(r+1)}.$$

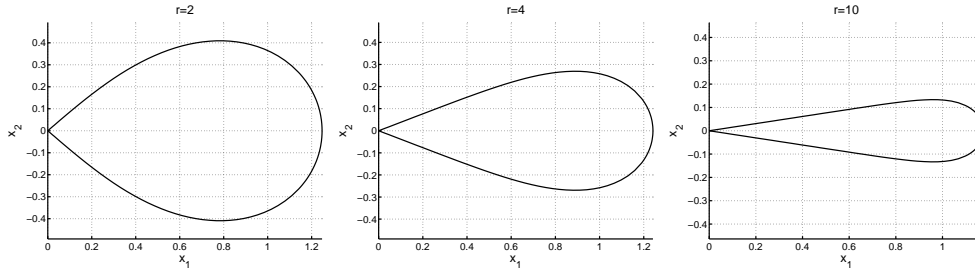


FIG. 4.1. Solution loops to problem(4.3) when $r = 2$ (left), $r = 4$ (middle), $r = 10$ (right).

5. Define a function f by $f(x_1) = g(\mu x_1)/\mu$, for $x_1 \in [0, a_0/\mu]$. It can be verified that the obtained f satisfies (4.9) and (4.10) for $\lambda_2/\lambda_3 = \kappa_0\mu^{-r}$ and $C = C_0\mu^{-r}$, and is indeed the desired function whose graph c follows during $[0, \frac{1}{2}]$. If one is interested only in the shape, not the scale, of the solution to problem (4.2), then the last step can be skipped.

Remark 6 *The time parameterization of c is recovered using the fact that c has constant speed along the graph of f on $[0, \frac{1}{2}]$. Or alternatively, by (4.6),*

$$t = \Phi(x_1) \triangleq \int_0^{x_1} \frac{dx_1}{-\lambda_3 \sqrt{C^2 - (\lambda_2/\lambda_3 + x_1^r)^2}}, \quad 0 \leq x_1 \leq a.$$

Note that $\Phi(x_1)$ is a strictly increasing function satisfying $\Phi(a) = \frac{1}{2}$. Once the function $\Phi(x_1)$ is determined, x_1 is determined as $x_1 = \Phi^{-1}(t)$.

Figure 4.1 plots the solution loops to problem (4.3) obtained from the above procedure for the case $r = 2, 4, 10$. In particular, in the Martinet distribution case ($r = 2$), the solution loop is part of the trajectory of a charged particle moving on \mathbb{R}^2 in a magnetic field \mathbf{B} with linear components and direction perpendicular to the plane (see the more complete plot of the motion in Fig 2.2 of the Alfvén's book [18, pp. 15]). The trajectory is called the *grad \mathbf{B} drift* since it exhibits an overall drift orthogonal to the direction $\text{grad}|\mathbf{B}|$. Indeed, much more has been known for the Martinet distribution: for example, its unit sub-riemannian sphere around the origin has been characterized and plotted using elliptic integrals in [21]. That the equation governing the solution loops is integrable in the $r = 2$ case has also been pointed out in [23] previously.

5. A Three-Segment Linked-Mass System. We now return to the motivating example in Section 1, and consider the case with $n = 2$. So the linked-mass system consists of four nodes, and three links whose orientations are given by the angles θ_i , $i = 1, 2, 3$. The configuration space is \mathbb{T}^3 , with a riemannian metric given by (1.6), where

$$\Delta = (\Delta_{ij})_{i,j=1}^3 = \frac{1}{4} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

By (1.8), the co-dimension one distribution \mathcal{H} is the kernel of the one-form

$$\omega = \sum_{i,j=1}^3 \Delta_{ij} \cos(\theta_i - \theta_j) d\theta_j.$$

Suppose now that the linked-mass system is at the initial configuration q corresponding to $\theta_1 = \theta_2 = \theta_3 = 0$, i.e. the three segments of the system are all aligned in the positive horizontal direction. To compute the rank at q , we perform the following coordinate transformation in a neighborhood of q :

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{5}}{3} & \frac{2\sqrt{5}}{3} & -\frac{\sqrt{5}}{3} \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}, \quad \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{5}}{10} & -\frac{1}{2} & \frac{1}{3} \\ \frac{\sqrt{5}}{5} & 0 & \frac{1}{3} \\ -\frac{\sqrt{5}}{10} & \frac{1}{2} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}. \quad (5.1)$$

The choice of such a transformation serves several purposes. First, $\phi_3 = \theta_1 + \theta_2 + \theta_3$ is the direction along the fibers of \mathbb{T}^3 under the action of \mathbb{T} as described in Section 1. Second, the plane Π spanned by the ϕ_1 and ϕ_2 axes is transversal to the ϕ_3 axis, hence can be regarded as the shape space, at least locally around the origin. Third, the projection $d\pi : \mathcal{H}_q \rightarrow T_{(0,0)}\Pi$ is an isometry if Π is equipped with the canonical Euclidean metric with respect to the coordinates (ϕ_1, ϕ_2) . So no further change of coordinates within Π is needed.

In the new coordinates, q corresponds to the origin $\phi_1 = \phi_2 = \phi_3 = 0$, and

$$\begin{aligned} 4\omega &= 3d\theta_1 + 2C_{12}d\theta_2 + C_{13}d\theta_3 + 2C_{12}d\theta_1 + 4d\theta_2 + 2C_{23}d\theta_3 + C_{13}d\theta_1 + 2C_{23}d\theta_2 + 3d\theta_3 \\ &= \frac{\sqrt{5}}{5}(1 + C_{12} + C_{23} - C_{13})d\phi_1 + (C_{23} - C_{12})d\phi_2 + \frac{2}{3}(5 + 2C_{12} + 2C_{23} + C_{13})d\phi_3, \end{aligned}$$

where C_{12}, C_{23}, C_{13} are defined by

$$\begin{aligned} C_{12} &\triangleq \cos(\theta_1 - \theta_2) = \cos\left(\frac{3\sqrt{5}}{10}\phi_1 + \frac{1}{2}\phi_2\right), \\ C_{23} &\triangleq \cos(\theta_2 - \theta_3) = \cos\left(\frac{3\sqrt{5}}{10}\phi_1 - \frac{1}{2}\phi_2\right), \\ C_{13} &\triangleq \cos(\theta_1 - \theta_3) = \cos\phi_2. \end{aligned}$$

From the above equations, the kernel of ω is the same as the kernel of

$$\Theta = -\frac{3(C_{12} - C_{23})}{2(5 + 2C_{12} + 2C_{23} + C_{13})}d\phi_1 + \frac{3(1 + C_{12} + C_{23} - C_{13})}{2\sqrt{5}(5 + 2C_{12} + 2C_{23} + C_{13})}d\phi_2 + d\phi_3,$$

which is of the standard form (2.1). Note that C_{12}, C_{23}, C_{13} are independent of ϕ_3 , as the distribution \mathcal{H} is invariant to the bundle structure $\pi : (\phi_1, \phi_2, \phi_3) \mapsto (\phi_1, \phi_2)$.

The form α defined in (2.4) is given by

$$\begin{aligned} \alpha &= -\frac{3(C_{12} - C_{23})}{2(5 + 2C_{12} + 2C_{23} + C_{13})}d\phi_1 + \frac{3(1 + C_{12} + C_{23} - C_{13})}{2\sqrt{5}(5 + 2C_{12} + 2C_{23} + C_{13})}d\phi_2 \\ &= \frac{3 \sin\left(\frac{3\sqrt{5}}{10}\phi_1\right) \sin\left(\frac{1}{2}\phi_2\right)}{5 + 4 \cos\left(\frac{3\sqrt{5}}{10}\phi_1\right) \cos\left(\frac{1}{2}\phi_2\right) + \cos\phi_2}d\phi_1 \\ &\quad + \frac{3(1 + 2 \cos\left(\frac{3\sqrt{5}}{10}\phi_1\right) \cos\left(\frac{1}{2}\phi_2\right) - \cos\phi_2)}{2\sqrt{5}(5 + 4 \cos\left(\frac{3\sqrt{5}}{10}\phi_1\right) \cos\left(\frac{1}{2}\phi_2\right) + \cos\phi_2)}d\phi_2, \end{aligned}$$

and $\beta = d\alpha = f(\phi_1, \phi_2)d\phi_1 \wedge d\phi_2$, where $f(\phi_1, \phi_2)$ is given by

$$f(\phi_1, \phi_2) = \frac{-3 \sin(\frac{3\sqrt{5}}{10}\phi_1)}{5(5 + 4 \cos(\frac{3\sqrt{5}}{10}\phi_1) \cos(\frac{1}{2}\phi_2) + \cos \phi_2)} \left\{ 4 \cos(\frac{1}{2}\phi_2) + \frac{\cos(\frac{3\sqrt{5}}{10}\phi_1)[10 \sin^2(\frac{1}{2}\phi_2) - 6 \cos^2(\frac{1}{2}\phi_2)] + 4 \sin^2(\frac{1}{2}\phi_2) \cos(\frac{1}{2}\phi_2)}{5 + 4 \cos(\frac{3\sqrt{5}}{10}\phi_1) \cos(\frac{1}{2}\phi_2) + \cos \phi_2} \right\}. \quad (5.2)$$

One can verify that

$$f(0, 0) = 0, \quad \frac{\partial f}{\partial \phi_1}(0, 0) = -\frac{51}{250} \neq 0, \quad \frac{\partial f}{\partial \phi_2}(0, 0) = 0.$$

As a result of Lemma 2, the rank at q is three.

Suppose that the test function $\mathcal{F} : \mathbb{R} \rightarrow \mathbb{R}$ is the identity map. To find the asymptotically optimal loop c based at q in the plane Π solving Problem 1, we can replace β by its first order approximate $-\frac{51}{250}\phi_1 d\phi_1 \wedge d\phi_2$, which is exactly of the form (4.1) with $r = 2$. Thus the results in Section 4 can be applied directly here.

In particular, the optimal loop c is computed in Section 4, and plotted in Figure 4.1 with coordinates $x_1 = \phi_1$ and $x_2 = \phi_2$. Horizontally lifting c in Π to a curve γ based at q in the (ϕ_1, ϕ_2, ϕ_3) coordinates, transforming γ back to the $(\theta_1, \theta_2, \theta_3)$ coordinates using transformation (5.1), and finally using equations (1.4) and (1.5), we obtain an asymptotically most efficient motions for the linked-mass system to turn starting from the initially aligned position. Figure 5.1 shows the snapshots of the motions obtained numerically at equally spaced time instances. Note that a relatively large scale of c is chosen in the plots to make this asymptotic motion more obvious.

Remark 7 *By equation (5.2), in a neighborhood of the origin in the (ϕ_1, ϕ_2) coordinates, $\beta = 0$ if and only if $\phi_1 = 0$, i.e., if and only if $\theta_1 + \theta_3 - 2\theta_2 = 0$. The rank is three at points satisfying this condition and two otherwise. As a result, when the system starts from a shape close to the aligned one, in the asymptotic sense it is more difficult to turn if its initial position is such that $\theta_1 + \theta_3 - 2\theta_2 = 0$.*

6. Extension to Principal Bundles. Due to their asymptotic nature, our results can be easily generalized to more complicated spaces such as principal bundles, as is described briefly in the following. For a Lie group G with Lie algebra \mathfrak{g} , a principal G -bundle $\pi : Q \rightarrow M$ is a fiber bundle whose structural group G acts freely and transitively on each fiber from the right. So each fiber is a copy of G and the vertical space $\mathcal{V}_q = \text{im}(\sigma_q)$ at each $q \in Q$ can be identified with \mathfrak{g} via the map $\sigma_q : \xi \in \mathfrak{g} \mapsto q \cdot \xi \in T_q Q$. A connection form Θ on Q is a \mathfrak{g} -valued one-form with $\ker \Theta_q \oplus \mathcal{V}_q = T_q Q$ and $\Theta_q \circ \sigma_q = \text{id}_{\mathfrak{g}}$, $\forall q \in Q$. Thus $\mathcal{H} \triangleq \ker \Theta$ defines a horizontal distribution on Q invariant under the action of G . The holonomy $h(c)$ of a loop c in M based at m can be identified as an element of G , and is determined up to a conjugacy class in G when varying the base point q of the horizontal lift [10].

Suppose that $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is a sub-riemannian metric on \mathcal{H} invariant under the action of G . Such $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is obtained by lifting a riemannian metric $\langle \cdot, \cdot \rangle_M$ on M . For a loop c in M based at m , we can define the rank $r_q(m) \in \mathbb{N}$ and the asymptotic holonomy $\eta_q(c) \in \mathfrak{g}$ such that $h(c_\epsilon) \sim \eta_q(c)[L(c_\epsilon)]^{r_q(m)}$ as $\epsilon \rightarrow 0$. Here the scaled loop c_ϵ is defined by identifying M locally around m with an open subset of \mathbb{R}^n via a coordinate map, for example, the inverse of the exponential map $\exp : T_m M \rightarrow M$. It

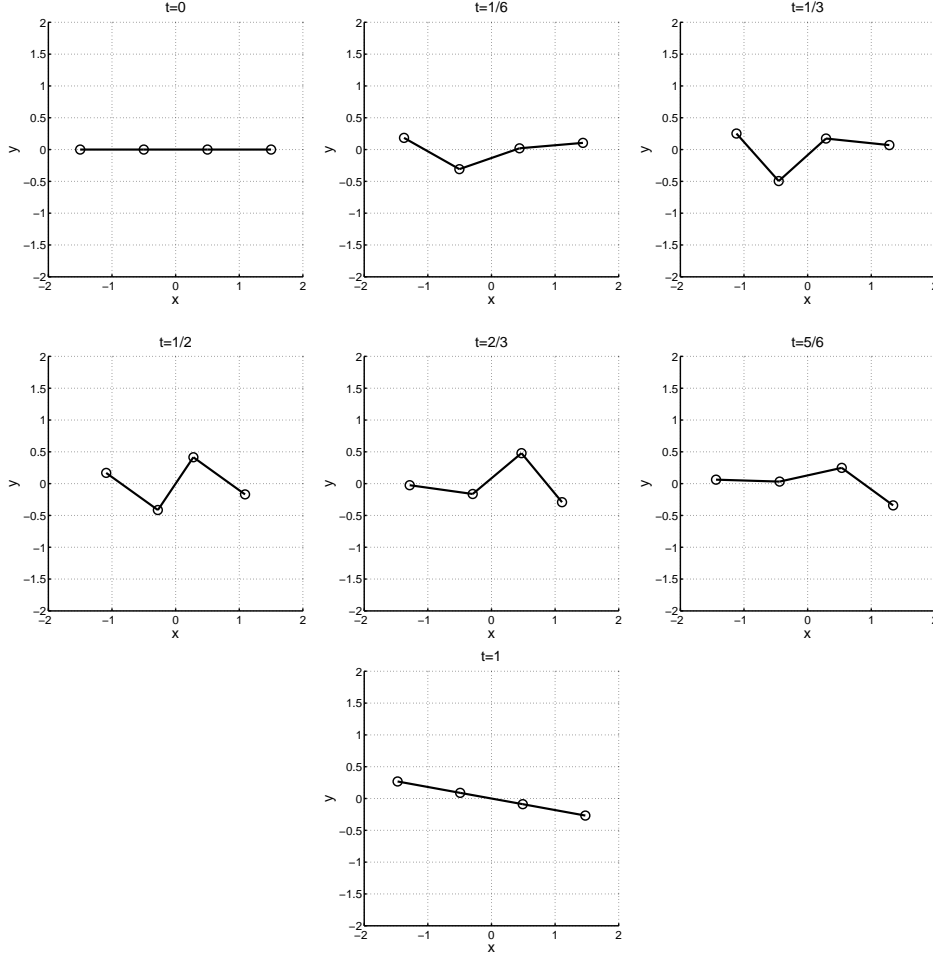


FIG. 5.1. Snap shots of the linked-mass system turning.

is easy to see that both $r_q(m)$ and $\eta_q(c)$ are independent of the choice of the coordinate maps; thus they are well defined. A test function is a map $\mathcal{F} : \mathfrak{g} \rightarrow \mathbb{R}$. For example, \mathcal{F} can be the inertial tensor $\mathcal{F}(\xi) = \langle \sigma_q(\xi), \sigma_q(\xi) \rangle_Q$, $\forall \xi \in \mathfrak{g}$, for some metric $\langle \cdot, \cdot \rangle_Q$ on Q that restricts to $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ on \mathcal{H} .

With the above notions, we can define the asymptotic holonomic efficiency problem on principal bundles. In a neighborhood of q , Q and $\mathbb{R}^n \oplus \mathfrak{g}$ are the same in terms of computing asymptotic holonomy. For example, when the distribution is of rank two at a point $m \in M$, it can be proved similarly as in Theorem 1 that the optimal loop c is a circle in a plane spanned by two tangent vectors u_m, v_m in $T_m M$. Finding the optimal circle then becomes finding these two tangent vectors u_m and v_m that maximizes a sectional curvature-like term $\mathcal{F}[\beta(u_m, v_m)]$ as in (3.4). In higher rank case, however, finding the optimal solutions is a much more challenging problem.

Finally, in terms of infinitesimal deformations, the general case when G is non-abelian looks exactly the same as the abelian case locally. Therefore, our results can be extended to the general non-abelian case without added difficulty [3].

Acknowledgment: The authors would like to thank Richard Montgomery for

pointing out some of the references, and Alan Weinstein for helpful discussions. We are also grateful for the anonymous reviewers for their constructive comments and for pointing out the many links to the existing literature.

Appendix A. Computation of the Metric on \mathbb{T}^{n+1} .

Plugging equation (1.4) into (1.5), we have

$$q_i = \frac{1}{n+2} \sum_{j=1}^{i-1} j(\cos \theta_j, \sin \theta_j)^T - \frac{1}{n+2} \sum_{j=i}^{n+1} (n+2-j)(\cos \theta_j, \sin \theta_j)^T. \quad (\text{A.1})$$

As a result, the tangent vector $\frac{\partial}{\partial \theta_l}$ at each point of \mathbb{T}^{n+1} is pushed forward by the embedding defined by (1.4) and (1.5) to a tangent vector in \mathbb{R}^{2n+4} :

$$\frac{\partial q}{\partial \theta_l} = \left[\underbrace{\frac{-n+2-l}{n+2}(-\sin \theta_l, \cos \theta_l)}_{\text{repeated } l \text{ times}} \quad \underbrace{\frac{l}{n+2}(-\sin \theta_l, \cos \theta_l)}_{\text{repeated } n+2-l \text{ times}} \right]^T \in \mathbb{R}^{2n+4}. \quad (\text{A.2})$$

The metric on \mathbb{T}^{n+1} can be derived from the standard metric on \mathbb{R}^{2n+4} as: $\langle \frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j} \rangle = \left(\frac{\partial q}{\partial \theta_i} \right)^T \cdot \left(\frac{\partial q}{\partial \theta_j} \right)$. Using (A.2), we can easily verify that $\langle \frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j} \rangle = \frac{i(n+2-j)}{n+2} \cos(\theta_i - \theta_j)$ if $i < j$; and $\langle \frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j} \rangle = \frac{(n+2-i)j}{n+2} \cos(\theta_i - \theta_j)$ if $i \geq j$.

REFERENCES

- [1] J. Hu, S. Simic, and S. Sastry. How should a snake turn on ice: A case study of the asymptotic isoholonomic problem. In *Proc. IEEE Int. Conf. on Decision and Control*, volume 3, pages 2908–2913, Maui, HI, Dec. 2003.
- [2] T. R. Kane and M. P. Scher. A dynamical explanation of the falling cat phenomenon. *International J. Solids and Structures*, 5(7):663–670, 1969.
- [3] A. Shapere and F. Wilczek. Self-propulsion at low Reynolds number. *Phys. Rev. Lett.*, 58(20):2051–2054, 1987.
- [4] R. Montgomery. The isoholonomic problem and some applications. *Comm. Math. Phys.*, 128(3):565–592, 1990.
- [5] J. M. Lighthill. On the squirming motion of nearly spherical deformable bodies liquids at very small reynold number. *Comm. Pure Appl. Maths*, 5(2):109–118, 1952.
- [6] S. Childress. *Mechanics of swimming and flying*. Cambridge University Press, 1981.
- [7] J. R. Blake. Self propulsion due to oscillations on the surface of a cylinder at low Reynolds numbers. *Bull. Australian Math. Soc.*, 5:255–264, 1971.
- [8] A. Shapere and F. Wilczek. Efficiencies of self-propulsion at low Reynolds number. *J. Fluid Mechanics*, 198:587–599, 1989.
- [9] J. Koiller and J. Delgado. On efficiency calculations for nonholonomic locomotion problems: an application to microswimming. *Reports on Mathematical Physics*, 42(1):165–183, 1998.
- [10] R. Montgomery. *A Tour of Subriemannian Geometries, Their Geodesics and Applications*. Amer. Math. Society, Providence, RI, 2002.
- [11] A. M. Bloch. *Nonholonomic Mechanics and Control*. Springer-Verlag, 2003.
- [12] J. B. Baillieul. Geometric methods for nonlinear optimal control problems. *J. Optimization Theory and Application*, 25(4):517–548, 1978.
- [13] R. W. Brockett. Control theory and singular Riemannian geometry. In P. J. Hilton and G. S Young, editors, *New Directions in Applied Mathematics*, pages 11–27. Springer-Verlag, 1982.
- [14] S. Sastry and R. Montgomery. The structure of optimal controls for a steering problem. In *Proc. 2nd IFAC Symposium on Nonlinear Control System Design (NOLCOS'92)*, pages 385–390, Bordeaux, France, 1992.
- [15] J Cortes, S. Martinez, J. P. Ostrowski, and K. A. McIsaac. Optimal gaits for dynamic robotic locomotion. *Int. J. Robotics Research*, 20(9):707–728, 2001.
- [16] A. Guichardet. On rotation and vibration motions of molecules. *Annales de l'institut Henri Poincaré (A) Physique theorique*, 40(3):329–342, 1984.

- [17] J. E. Marsden and T. S. Ratiu. *Introduction to Mechanics and Symmetry*. Springer-Verlag, New York, NY, 2nd edition, 1999.
- [18] H. Alfven. *Cosmical Electrodynamics*. Oxford University Press, 1950.
- [19] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko. *Mathematical Theory of Optimal Processes*. Interscience Publishers, 1963.
- [20] L. de Lathauwer, B. de Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Control and Optimization*, 21(4):1253–1278, 2000.
- [21] A. A. Agrachev, B. Bonnard, M. Chyba, and I. Kupka. Sub-riemannian sphere in martinet flat case. *J. ESAIM: Control, Optimisation and Calculus of Variations*, 2:377–448, 1997.
- [22] W. Liu and H. J. Sussmann. Shortest paths for sub-riemannian metrics on rank-two distributions. *Memoirs of the American Mathematical Society Series*, 118(564), 1995.
- [23] R. Yang. *Nonholonomic Geometry, Mechanics and Control*. PhD thesis, University of Maryland, College Park, MD, August 1992.