Exploring the link between Geodesically Convex Optimization and Contraction Analysis

Mohamad Qadri and Chiheb Boussema

1 Introduction

Optimization plays an important role in several fields, including robotics control and machine learning. An important subfield of optimization is convex optimization which enjoys several favorable properties and mature solution methods. Because of its limited definition, the scope of convex optimization may remain restrained and inapplicable to several complex robotic situations such as legged locomotion. Fortunately though, in some instances, problems that are not convex in a Euclidean sense can be recast over a Riemannian manifold with a suitable Riemannian metric such that they become "geodesically convex along shortest paths on this manifold". This extended notion of convexity called geodesic convexity [1] then turns hard-to-analyze and solve problems into ones that again enjoy favorable properties.

Interestingly, there is another field that uses similar notions of metrics on Riemannian manifolds and geodesic shortest paths: contraction theory [2]. The latter has been surging as a powerful and unifying tool for the stability analysis of nonlinear non-autonomous dynamical systems. Within this framework, the existence of a suitable Riemannian metric that contracts small displacements along the system's flow implies the exponential convergence of system trajectories to each other.

Given the potential importance of both fields and their use of similar notions, one can expect to find connections between both. In this study therefore, we proposed to (a) explore notions related to the fields of geodesic convexity and contraction theory, (b) the link between both as presented in a recent paper by Wensing and Slotine [3], and (c) some of its implications on an optimization problem.

2 Differential Geometry

We started by studying topics in topology and differential geometry that we deemed required to gain a good understanding of the field of geodesic convexity. We mostly used [4], [5] and [6] for the following.

2.1 Differentiable Manifolds

The following definitions are based on Frederic P Schuller's lectures on General Relativity (cite). We start our building blocks from the notion of a set and give a definition of a topological space:

Definition 1 *Let M be a set and $P(M)$ be the power set of M. A set $O \subseteq P(M)$ is called a topology, if it satisfies the following properties:*

- $\varnothing \in O, M \in O$
- $U \in O, V \in O \longrightarrow U \cap V \in O$
- $U_i \in O \longrightarrow \cup U_i \in O$

where $\varnothing$ is the empty set and U, V, U_i are sets in $P(M)$. The tuple (M, O) is a topological space.

[4]

Definition 2 *A topological space (M, O) is called a d-dimensional topological manifold if $\forall p \in M : \exists U \in O : p \in U, \exists x : U \longrightarrow x(U) \subseteq \mathbb{R}^d$ satisfying the following:*

- *x is invertible: $x^{-1} : x(U) \longrightarrow U$*
- *x^{-1} is continuous*
- *x is continuous wrt (M, O) and $(\mathbb{R}^d, O_{std})$*

where O_{std} is the standard topology on $\mathbb{R}^d$

- *The tuple (U, x) is a chart of (M, O)*
- *$x : U \longrightarrow x(U) \in \mathbb{R}^d$ is called the chart map*

[4]

Let 2 intersecting region U_α and U_β on a manifold M and let ψ_α and ψ_β be maps from U_α to $\mathbb{R}^d$ and U_β to $\mathbb{R}^d$ respectively. The map $\psi_{\alpha\beta} = \psi_\beta \circ \psi_\alpha^{-1}$ is called a chart transition map which maps an open set of $\mathbb{R}^d$ to another open set of $\mathbb{R}^d$

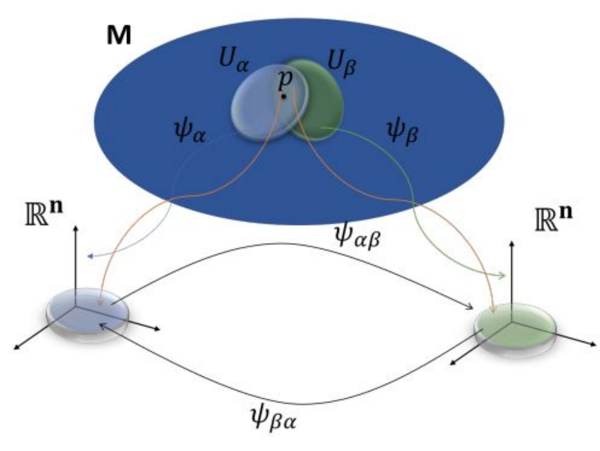

Figure 1: Transition maps (figure borrowed from [6])

The manifold M is called a differentiable manifold if all transition maps are differentiable. More generally, if all transition functions are k-times differentiable, then M is a C^k-manifold. If all transition functions have derivatives of all orders, then M is said to be a C^∞ manifold or smooth manifold. [6] The idea is to disregard all charts where the transition maps are not differentiable so we only keep compatible charts while making sure that we have enough charts to map the entire manifold.

2.2 Riemannian Manifolds

2.2.1 Metric tensors

We established the notion of differentiability on a manifold. However, we still lack the means of measuring angles and distances which are essential in the theory of optimization. Endowing a smooth manifold with a metric tensor allows to do just that: it gives us a way to measure the distance of a geodesic between 2 points on the manifold.

Definition 3 *More formally, let M be a smooth manifold. At each point p, we have a vector space T_pM which is the collection of tangent vectors at point p. Define a function g_p at point p that maps 2 tangent vectors at point p to $\mathbb{R}$ ($g_p : T_pM \times T_pM \longrightarrow \mathbb{R}$). If the the function g_p is* ***linear, symmetric and non-degenerate*** *then g_p is a metric tensor at point p*

We refer to [6] for the definition of the bold terms in the definition.

Definition 4 *Let M be a smooth manifold and g be a metric tensor over M. If for any open set U and smooth vector fields X and Y over U, the function $g(X,Y)[p] := g_p(X_p, Y_p)$ is a smooth function of p, and g is positive definite, then (M, g) is called a* ***Riemannian manifold***

[6]

2.3 The Riemannian Hessian and the Christoffel Symbols

One of the first definitions we encountered in [3] was the definition of geodesic strong convexity which depend on the Riemannian Hessian and on the Christoffel Symbols of the second kind. We studied areas in differential geometry to gain an understanding of these 2 concepts.

2.3.1 How to find geodesics in curved space? What are Christoffel Symbols?

Geodesics are shortest path between 2 points with respect to a metric. More precisely: *"A geodesic is defined as the curve such that a point moving along the curve with the velocity of constant magnitude (i.e. the velocity can change its direction but not its magnitude) has the acceleration vector perpendicular to the given surface, i.e.*

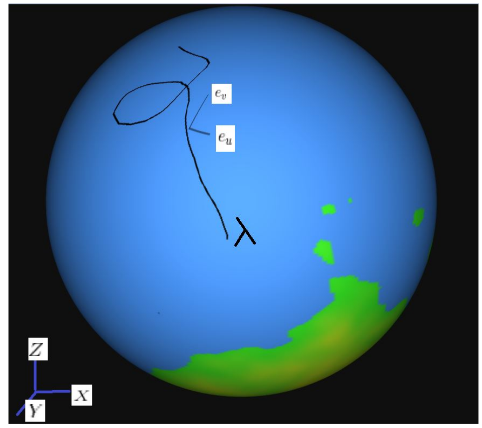

Figure 2: path λ on a unit sphere with intrinsics and extrinsics coordinates

the acceleration component tangent to the given surface is zero." [7]. We start with the simple example of a 2D sphere and then generalize the derivation to any surface. Let's consider the situation in figure 1 where we show a parametric path λ on a unit 2D sphere. We have an intrinsic coordinate system $I_b = (u, v)$ and a fixed extrinsic coordinate system $E_b = (X, Y, Z)$. We also let a position vector $R(u, v) = (X(u, v), Y(u, v), Z(u, v))$. The relation between between E_b and I_b is:

$$X = cos(u)sin(v)$$
$$Y = sin(v)sin(u)$$
$$Z = cos(u)$$

The tangent vectors to the curve parametrized by λ are given by $\frac{dR}{d\lambda} = \frac{dR}{du}\frac{du}{d\lambda} + \frac{dR}{dv}\frac{dv}{d\lambda}$. The intrinsic basis vectors are given by:

$$\frac{dR}{du} = e_u \quad \frac{dR}{dv} = e_v. \tag{1}$$

The acceleration vectors to the curve λ is given by the derivative of the velocity:

$$\frac{d}{d\lambda}\frac{dR}{d\lambda} = \frac{d^2R}{dud\lambda}\frac{du}{d\lambda} + \frac{du^2}{d\lambda^2}\frac{dR}{du} + \frac{d^2R}{dvd\lambda}\frac{dv}{d\lambda} + \frac{dv^2}{d\lambda^2}\frac{dR}{dv}$$

Using chain rule and reorganizing, we get the following equation for the acceleration vectors:

$$\frac{d^2R}{d\lambda^2} = \frac{d^2u}{d\lambda^2}\frac{\partial R}{\partial u} + \frac{du}{d\lambda}\left(\frac{du}{d\lambda}\frac{\partial^2 R}{\partial u^2} + \frac{dv}{d\lambda}\frac{\partial^2 R}{\partial u \partial v}\right) + \frac{d^2v}{d\lambda^2}\frac{\partial R}{\partial v} + \frac{dv}{d\lambda}\left(\frac{du}{d\lambda}\frac{\partial^2 R}{\partial v^2}\frac{du}{d\lambda}\frac{\partial^2 R}{\partial u \partial v}\right) \tag{2}$$

Given (1), we know that the first and third term in the equation above are tangential components. However, we cannot make any conclusion regarding the second and fourth term. We can re-write equation(2) in Einstein's notation (let $u^1 = u$ and $u^2 = v$):

$$\frac{dR^2}{d\lambda^2} = \frac{dR}{\partial u^i}\frac{d^2u^i}{d\lambda^2} + \frac{du^i}{\partial\lambda}\frac{du^j}{\partial\lambda}\frac{\partial^2 R}{\partial u^i \partial u^j} \tag{3}$$

The second term are vectors that we can express in terms of u^1, u^2 and $u^1 \times u^2$ (can write them as a linear combination of $\frac{\partial R}{\partial u^1}, \frac{\partial R}{\partial u^2}, n)$ where n is normal to the surface.

$$\frac{\partial^2 R}{\partial u^i u^j} = \Gamma^1_{ij}\frac{\partial R}{\partial u} + \Gamma^2_{ij}\frac{\partial R}{\partial u^2} + L_{ij}n \tag{4}$$

Use Einstein's notation again to sum over the Γ^k_{ij} to get:

$$\frac{\partial^2 R}{\partial u^i u^j} = \Gamma^k_{ij}\frac{\partial R}{\partial u^k} + L_{ij}n \tag{5}$$

The **Christoffel symbols** Γ^k_{ij} are the coefficients of the tangential components. The **second fundamental form** L_{ij} give us the normal component of $\frac{\partial^2 R}{\partial u^i u^j}$. We note that these definitions depend on embedding the surface into a higher dimentional space (we relied on extrinsic coordinates in the derivation). We will later derive the equation of the Christoffel symbols using only intrisincs coordinates which will give us the equations of the Christoffel symbols of the second kind used in the definition of the Riemannian Hessian.

To recover an equation for the Christoffel symbols, We multiply (5) by $\frac{\partial R}{\partial u^l}$. Since $\frac{\partial R}{\partial u^l}$ is orthogonal to n, the dot product involving n is equal to 0. We get:

$$\frac{\partial^2 R}{\partial u^i u^j}\frac{\partial R}{\partial u^l} = \Gamma^k_{ij}\frac{\partial R}{\partial u^k}\frac{\partial R}{\partial u^l}$$

Since $\frac{\partial R}{\partial u^k}\frac{\partial R}{\partial u^l} = g_{kl}$ where g_{kl} is the metric tensor, we get:

$$\frac{\partial^2 R}{\partial u^i u^j}\frac{\partial R}{\partial u^l} = \Gamma^k_{ij} g_{kl}$$

Multiply by the inverse metric tensor on both side. Since $g_{kl}g^{lm} = \delta^m_k$ where δ^m_k is the Kronecker delta, We get an equation for the christoffel symbols: $\Gamma^m_{ij} = \frac{\partial^2 R}{\partial u^i u^j}\frac{\partial R}{\partial u^l} g^{lm}$. (We use the Kronecker delta property $\Gamma^k_{ij}\delta^m_k = \Gamma^m_{ij}$) Combining this result with equation 6 (change variable m with k), a geodesic curve is a solution to the geodesic equation:

$$\frac{d^2 u^k}{d\lambda^2} + \Gamma^k_{ij}\frac{du^i}{d\lambda}\frac{du^j}{d\lambda} = 0 \tag{6}$$

We verified this result on the flat plane and the sphere which we will omit here.

2.3.2 Differentiation on Manifolds and Covariant derivatives

Our goal is to optimize a function defined on a manifold. Optimizing a function defined on a Euclidean space eventually involves moving in the direction of the negative gradient of the function f. This means we need a way to compare the rate change of a function (or a general vector field) with respect to another vector field. *Differentiation inherently requires comparing the change in a function or vector value when we move from one point to another close by point (in a given direction)* [6]

The Covariant derivatives is a way of specifying a derivative along tangent vectors of a manifold.

We first studied covariant derivatives of a vector field from an extrinsic view. Citing wikipedia: *In the special case of a manifold isometrically embedded into a higher-dimensional Euclidean space, the covariant derivative can be viewed as the orthogonal projection of the Euclidean directional derivative onto the manifold's tangent space.* Therefore we can view the covarient derivative $\nabla_w v$ as the rate of change of a vector field v in a direction w with the normal component substracted. We can derive an equation for the covariant derivate as follows:
Given a vector field v. We express v in terms of its components with respect to the basis e_j. The derivative of v with respect to a direction w can also be expressed in terms of the basis e_j (we note that this basis is not necessary constant at each point)

$$\frac{\partial}{\partial w}v = \frac{\partial}{\partial w}(v^j e_j)$$
$$\frac{\partial}{\partial w}v = \frac{\partial v^j}{\partial w}e_j + v^j\frac{\partial e_j}{\partial w}$$

$\frac{\partial e_j}{\partial w}$ can be written as a linear combination of the tangent basis and the christoffel symbols:

$$\frac{\partial e_j}{\partial w} = \Gamma^1_{ij}e_1 + \Gamma^2_{ij}e_2 = \Gamma^k_{ij}e_k$$

and we obtain an equation of the covariant derivative:

$$\nabla_w v = (\frac{\partial v^k}{\partial w} + v^j\Gamma^k_{ij})e_k \tag{7}$$

2.4 An intrinsic view of the Covariant Derivative

Up to this point, we relied on an extrinsic coordinate system to study covariant derivatives and to retrieve the equations of the Christoffel symbols. We now study these concepts relying only on the intrinsic coordinates (for example e_u and e_v in figure 1). This means that we need to abandon the idea of a global constant coordinate (XYZ coordinate in figure 1), global origin and position vectors. We also ignore all vector components that are perpendicular to the surface. From the intrinsic point of view, we require that $e_i.e_j$ is given. This means that we need a metric tensor to be provided to us. Equation (7) of the covariant derivative remains valid. However, the equation of the Christoffel symbols will change (since we previously relied on the dot product and on the position vector R).

We introduce briefly the concept of a **connection** and of **parallel transport**. Given a tangent vector on T_pM, the covariant derivatives gives us a way to parallel transport a vector along a curve to a different tangent space T_qM of a different point q such that the vector stays "as constant as possible". It is a way to provide a connection between tangent spaces in a curved space. We note that it is not possible to define a constant vector field on a curved surface. Finally, we note that there exists many possible Covariant derivatives, to get a unique solution, we will require our solution to satisfy the following 2 properties:

2.4.1 Metric Compatibility property

$$\nabla_w(v.u) = (\nabla_w v).u + v.(\nabla_w w) \tag{8}$$

2.4.2 Torsion free

$$\begin{aligned}\nabla_w u - \nabla_v w = [v, w] = 0 \\ \nabla_w u = \nabla_v w\end{aligned} \tag{9}$$

where [v, w] is the lie bracket. From the torsion free property, we derive the following properties:

$$\begin{aligned}\nabla_{e_j} e_i = \nabla_{e_i} e_j \\ \frac{\partial}{\partial u^i}\frac{\partial}{\partial u^j} = \frac{\partial}{\partial u^j}\frac{\partial}{\partial u^i} \\ \frac{\partial}{\partial u^i} e_j = \frac{\partial}{\partial u^j} e_i\end{aligned} \tag{10}$$

Since: $\nabla_{e_i} e_j = \Gamma^k_{ij} e_k$ **and** $\nabla_{e_j} e_i = \Gamma^k_{ji} e_k$ we get:

$$\Gamma^k_{ij} = \Gamma^k_{ji} \tag{11}$$

We are guaranteed the existence of such a connection by the **Fundamental Theorem of Riemannian Geometry**: On a Riemannian manifold, there is a unique connection which is torsion-free and compatible with the metric. This connection is called the **Levi-Civita connection**.

We now derive the equations of the Christoffel symbols for the Levi-Civita connection. The metric tensor is equal to $g_{ij} = e_i.e_j$. We take the covariant derivative of the metric tensor in the u_k component

$$\frac{\partial}{\partial u^k} g_{ij} = \frac{\partial}{\partial u^k}(e_i.e_j)$$

- Use the metric compatibility property

$$\frac{\partial}{\partial u^k}(e_i.e_j) = \frac{\partial e_i}{\partial u^k} e_j + e_i \frac{\partial e_j}{\partial u^k}$$

- express the partial derivatives as a linear combination of basis vectors

$$\frac{\partial}{\partial u^k} g_{ij} = (\Gamma^l_{ik} e_l) e_j + e_i(\Gamma^l_{jk} e_l)$$

$$\frac{\partial}{\partial u^k} g_{ij} = \Gamma^l_{ik}(e_l.e_j) + \Gamma^l_{jk}(e_i.e_l)$$

$$\frac{\partial}{\partial u^k} g_{ij} = \Gamma^l_{ik} g_{lj} + \Gamma^l_{jk} g_{il}$$

We write the formula above for three different indices:

$$\begin{aligned}\frac{\partial g_{ij}}{\partial u^k} &= \Gamma^l_{ik} g_{jl} + \Gamma^l_{jk} g_{il} \\ \frac{\partial g_{ki}}{\partial u^j} &= \Gamma^l_{kl} g_{jl} + \Gamma^l_{ij} g_{kl} \\ \frac{\partial g_{jk}}{\partial u^i} &= \Gamma^l_{ji} g_{kl} + \Gamma^l_{ki} g_{jl}\end{aligned}$$

Therefore:

$$\frac{\partial g_{ij}}{\partial u^k} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{jk}}{\partial u^i} = \Gamma^l_{ik} g_{jl} + \Gamma^l_{jk} g_{il} + \Gamma^l_{kl} g_{jl} + \Gamma^l_{ij} g_{kl} - \Gamma^l_{ji} g_{kl} - \Gamma^l_{ki} g_{jl}$$

Using (14), we can cancel out terms to get:

$$\frac{\partial g_{ij}}{\partial u^k} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{jk}}{\partial u^i} = 2\Gamma^l_{jk} g_{li}$$

Multiply by the inverse metric tensor :

$$g^{im}(\frac{\partial g_{ij}}{\partial u^k} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{jk}}{\partial u^i}) = 2\Gamma^l_{jk} g_{li} g^{im}$$

Using the property of the Kronocker delta and re-arraging terms:

$$\frac{1}{2} g^{im}(\frac{\partial g_{ij}}{\partial u^k} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{jk}}{\partial u^i}) = \Gamma^l_{jk} \delta^m_l$$

Swapping indices, we conclude that:

$$\boxed{\Gamma^k_{ij} = \frac{1}{2} g^{kl}(\frac{\partial g_{li}}{\partial u^j} + \frac{\partial g_{jl}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^l})}$$

2.5 Covariant derivative of covector fields

We can extend the notion of a covariant derivative to any Tensor field and more specifically to covector fields which is a (0, 1) tensor. We can easily show that the covarient derivative of a covector field α is:

$$\nabla_{\partial i}(\alpha) = (\frac{\partial \alpha_k}{\partial u^i} - \alpha_j \Gamma^j_{ik})\epsilon^k \tag{12}$$

where ϵ^k is the basis of the covector space. We omit the derivation here.

2.5.1 Cotangent spaces are covector fields

Let T_pM be the tangent space at point p on a manifold M, we define the cotangent space as the set of linear maps from the tangent space to R:

$$(T_pM)^* := \rho : T_pM \longrightarrow \mathbb{R}$$

The gradient of f at a point $p \in M$ is defined as:

$$\begin{aligned}(df)_p : T_pM \longrightarrow \mathbb{R} \\ X := (df)_p(X) := Xf\end{aligned}$$

Therefore $(df)_p \in TpM^*$. We define the components of the gradient with respect to the basis as:

$$((df)_p)_j = (\frac{\partial f}{\partial u^j})_p \tag{13}$$

2.5.2 The Riemmanian Hessian

Let (M,g) be a Riemannian manifold and ∇ its Levi-Civita connection. Let $f : M \longrightarrow R$ be a smooth function. We define the Hessian tensor as:

$$Hess(f) = \nabla\nabla f$$

Since f is a scalar field, it's first covariant derivative is its gradient:

$$Hess(f) = \nabla df \tag{14}$$

Since df_p is a covector, we use equation 12 (set $\alpha = df_p$) to get an equation for the components of the Riemannian Hessian (write df_p in terms of its components using 13):

$$\boxed{H_{ij} = (\frac{\partial^2 f}{\partial u^i \partial u^j} - \frac{\partial f}{\partial u^j}\Gamma^j_{ik})} \tag{15}$$

We note that we derived the equation of Christoffel Symbols by taking the derivative of the metric (page 4). Since in Euclidean flat space, the metric is constant at all points, the christoffel symbols will be all zeros. And so equation (15) reverts to the equation of the hessian matrix on Euclidean flat space.

3 Geodesic Convexity

We now give the definitions of geodesically convex sets and convex functions

3.0.1 Geodesically convex sets

Let (M, g) be a Riemannian manifold. A set $K \subseteq M$ is said to be totally convex with respect to g, if for any $p, q \in K$, any geodesic γ_{pq} that joins p to q lies entirely in K. [6]

3.0.2 Geodesically convex functions

Let (M, g) be a Riemannian manifold and $K \subseteq M$ be a totally convex set with respect to g. A function $f : K \longrightarrow R$ is said to be a geodesically convex function with respect to g if for any $p, q \in K$, and for any geodesic $\gamma_{pq} : [0, 1] \longrightarrow K$ that joins p to q,

$$\forall t \in [0,1] \quad f(\gamma_{pq}(t)) \leqslant (1-t)f(p) + tf(q).$$

[6]

3.0.3 g-strong convexity

Theorem 1 *A function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is said to be geodesically α-strongly convex (with $\alpha > 0$) in a symmetric positive definite metric M(x) if its Riemannian Hessian matrix H(x) satisfies:*

$$\mathbf{H}(x) \geq \alpha \mathbf{M}(x), \quad \forall x \in \mathbb{R}^n \tag{16}$$

The elements of the Riemannian Hessian are given as:

$$H_{ij} = \partial_{ij} f - \Gamma^k_{ij} \partial_k f \tag{17}$$

where $\partial_{ij} f = \frac{\partial^2 f}{\partial x_i \partial x_j}$ provide the elements of the conventional Euclidean Hessian and Γ^k_{ij} denotes the Christoffel Symbols of the second kind:

$$\Gamma^m_{ij} = \frac{1}{2}\sum_{k=1}^{n}[M^{mk}(\partial_j M_{ik} + \partial_i M_{jk} - \partial_k M_{ij})] \tag{18}$$

with $M^{ij} = (\mathbf{M^{-1}})_{ij}$. The function is g-convex when (1) holds with $\alpha = 0$

[3] Equation 17 has been derived in part section (2.5) and equation 18 in section (2.4)

3.0.4 Natural gradient

Consider $\mathbb{R}^n$ equipped with a Riemannian metric M(x). The natural gradient of the function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is the direction of steepest ascent on the manifold and is given by $M(x)^{-1}\partial_x f$ [3]

We conclude this section by summarizing the motivation of paper [3]: Given a non-convex function f defined on $\mathbb{R}^n$ (equipped with the standard metric), we can find a different positive semi-definite metric M such that f defined on the Riemannian manifold $(\mathbb{R}^n, M)$ is g-convex. We can then perform Natural Gradient Descent to find the minima (which is now the global minima) where the gradient descent paths are now the redefined geodesics.

4 Contraction Theory

Interestingly, Contraction Theory, a surging subfield of nonlinear dynamical systems analysis, also uses similar notions of metrics on Riemannian manifolds and geodesic shortest paths. To explore the link between it and geodesic convexity, we first explore basics of contraction. In this section we will mainly follow [8] to introduce contraction theory.

4.1 Introduction to Contraction

The analysis and control of nonlinear dynamical systems of the form

$$\dot{x} = f(t, x) \tag{19}$$

often relies on Lyapunov methods: either the linearization method which analyses stability questions around an equilibrium, or the direct method, which makes use of an energy-like scalar function to analyze the system stability [9]. Physical insight from the system in question, or its analogy to a physical system, is often used to craft this mechanical energy-like Lyapunov function, giving a sense of 'virtual' classical mechanics to this type of analysis.
By contrast, contraction theory takes a 'virtual fluids'

approach and analyses dynamical systems from a *differential* Lyapunov-like perspective. By investigating the dynamics of differential virtual displacements, it makes the remark that if one knows the linearization of a system everywhere, then one can actually infer quite a lot of information about the nonlinear system. This is in contrast to linearizing simply around equilibrium points. Indeed, contraction theory does not need the knowledge, nor the existence, of an attractor to conclude stability properties.
Intuitively, contraction analysis asks whether any 2 system trajectories (i.e., 2 particular solutions of (19)) converge to each other exponentially. In so doing, it separates questions of convergence and of final behavior. A contracting system therefore is a dynamical system (19) for which any 2 system trajectories converge to each other exponentially. This provides an interesting subtle alternative to error control in robotic applications. Indeed, instead of tracking an error function (e.g., $x(t) - x_{\text{des}}(t)$, where x_{des} is a desired trajectory), if we happen to know a particular trajectory of a contracting system, then we have a certificate that any other trajectory converges to the known trajectory exponentially. We will later give an example implementation of this interesting approach to the design of a reduced order observer for some states of a nonlinear underactuated dynamical system.

4.2 Contraction Basics

Consider the system (19), where the dimension of x is $\dim x = n$. Consider a volume, $\mathcal{V}$ of state space. Using Gauss Theorem, we have that $\frac{d}{dt}\mathcal{V} = \int_{\mathcal{V}} \operatorname{div}(\dot{x})$, where div is the divergence operator. Suppose $\operatorname{div}(\dot{x}) \leq -\lambda$, with $\lambda > 0$, then that means that $\frac{d}{dt}\mathcal{V} \leq -\lambda\mathcal{V}$, hence $0 \leq \mathcal{V} \leq \mathcal{V}(t=0)e^{-\lambda t}$. So the volume $\mathcal{V}$ tends to 0 exponentially. In other words, if the divergence of the dynamics of a system is negative, then any volume shrinks to 0. Is this enough to conclude that any system trajectory must converge somewhere? Unfortunately it isn't. Indeed the shrinkage of the volume to 0 only means that the system converges to a space of dimension $n-1$. We need something else to show that the volume tends to a (time-varying) point.

Now consider a virtual displacement δx, i.e, a differential variation at a fixed time. Consider the dynamics of the variation between two infinitesimally close trajectories x_1 and x_2 at any given time. We have $\delta x = x_2 - x_1$, so that

$$\frac{d}{dt}\delta x = \frac{\partial f}{\partial x}\delta x \tag{20}$$

Let us consider the variation of the squared norm $||\delta x||^2 = \delta x^T \delta x$:

$$\frac{d}{dt}\delta x^T \delta x = 2\delta x^T \frac{d}{dt}\delta x = 2\delta x^T \frac{\partial f}{\partial x}\delta x$$

Now suppose $\partial_x f \leq -\alpha I$[1], with $\alpha > 0$, then we have

$$\frac{d}{dt}\delta x^T \delta x \leq -2\alpha(\delta x^T \delta x)$$

Therefore, $||\delta x|| \to 0$ exponentially with rate α. Hence the two trajectories x_1 and x_2 converge to each other exponentially with rate α.
We can integrate the above reasoning, *à la Riemann*, for arbitrarily far trajectories. Consider x_1 and x_2 far away from each other at $t = 0$. Split the line connecting both of them into infinitesimally small sections. At time $t > 0$, as each of those segments will have shrunk exponentially (each virtual point may have travelled at different velocities), then the total length of the curve connecting $x_1(t)$ and $x_2(t)$ (no longer a straight line) will also have shrunk by $e^{-\alpha t}$. The straight line connecting $x_1(t)$ and $x_2(t)$ will be even shorter. Hence the analysis above is also valid for arbitrarily far trajectories.

Proposition 1 ([8, 2]) *Consider the system* $\dot{x} = f(t,x)$. *If* $\exists \alpha > 0, \forall x, \forall t, \frac{\partial f}{\partial x} \leq -\alpha I$, *i.e., all the eigenvalues of the symmetric part,* $\frac{1}{2}(\partial_x f + \partial_x f^T)$, *are strictly uniformly negative, then all system trajectories converge to each other exponentially with rate* α.

Example: Reduced order observer for an extended cart-pole system We apply Proposition 1 to design an exponentially converging observer on the angular states $\theta, \dot{\theta}$ of a cart-pole system extended to allow motion in the vertical direction z and with a damping term on the angular dynamics. The system dynamics are give by:

$$\begin{cases} \ddot{x} &= -\gamma\ddot{\theta}c_\theta + \gamma\dot{\theta}^2 s_\theta + \frac{u_x}{M} \\ \ddot{z} &= -\gamma\ddot{\theta}s_\theta - \gamma\dot{\theta}^2 c_\theta - g + \frac{u_z}{M} \\ \ddot{\theta} &= \frac{-1}{l}\ddot{x}c_\theta - \frac{1}{l}\ddot{z}s_\theta - \frac{g}{l}s_\theta - \frac{\nu}{l^2 m_b}\dot{\theta} \end{cases} \tag{21}$$

where the system states are $[x, z, \theta, \dot{x}, \dot{z}, \dot{\theta}]^T$, u_x and u_z are control inputs, c_θ (s_θ) signify $\cos\theta$ ($\sin\theta$), and all other variables are positive constants. We propose a reduced order observer to estimate $\theta, \dot{\theta}$: given a measured $\ddot{x}$ and $\ddot{z}$, consider the following observer dynamics:

$$\begin{cases} \dot{\hat{\theta}}_1 &= \hat{\theta}_2 \\ \dot{\hat{\theta}}_2 &= \frac{-1}{l}\ddot{x}c_{\hat{\theta}_1} - \frac{1}{l}(\ddot{z}+g)s_{\hat{\theta}_1} - \frac{\nu}{l^2 m_b}\hat{\theta}_2 \end{cases} \tag{22}$$

where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the estimators of θ and $\dot{\theta}$, respectively. The Jacobian of this system is:

$$J = \frac{\partial\dot{\hat{\theta}}}{\partial\hat{\theta}} = \begin{pmatrix} 0 & 1 \\ \phi & -b \end{pmatrix}$$

where $\phi = \frac{1}{l}\ddot{x}s_{\hat{\theta}_1} - \frac{1}{l}(\ddot{z}+g)c_{\hat{\theta}_1}$ and $b = \frac{\nu}{l^2 m_b}$. The eigenvalues of $(J + J^T)$ are $\lambda_{1,2} = -b \pm \sqrt{b^2 - (1+\phi)^2} < 0$,

[1] We note that for 2 symmetric matrices, $A \leq B$ means that $x^T A x \leq x^T B x$.

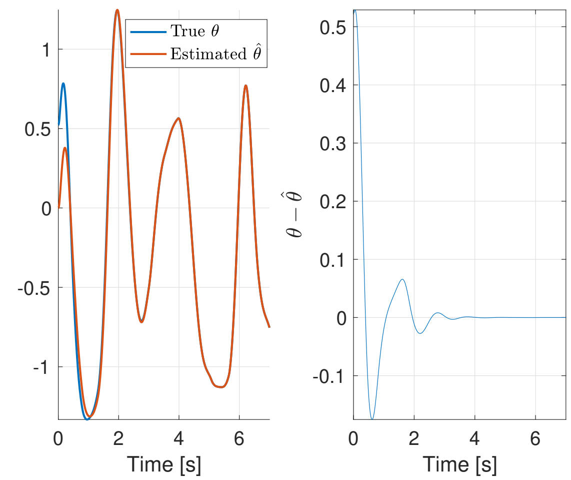

Figure 3: Convergence of the observer $\hat{\theta}, \dot{\hat{\theta}}$ to the true $\theta, \dot{\theta}$ even when starting with a random initial condition. (Left) Convergence of $\hat{\theta}$ to θ. (Right) Plot of the error $\theta - \hat{\theta}$ as a function of time.

therefore the observer dynamics (22) are contracting. Now since the true $\theta, \dot{\theta}$ is a solution to (22) (eq. 22 is in fact exactly the dynamics of θ when $\hat{\theta}_i$ is replaced by θ_i), then $\hat{\theta}, \dot{\hat{\theta}}$ converge to $\theta, \dot{\theta}$ exponentially. This can be confirmed in Fig. 3. Note how we used the subtle alternate approach we discussed in 4.1 where we first know the system is contracting, then we know a particular solution, from which we conclude that all trajectories must be converging to it.

While Proposition 1 provides a condition on dynamical systems to be contracting, it is in fact only a sufficient condition. A more general condition is derived hereafter. Consider the differential change of variables

$$\delta z = \underbrace{\frac{\partial z}{\partial x}(x,t)}_{\Theta(x,t)}\,\delta x \tag{23}$$

Then we have

$$\begin{aligned}\frac{d}{dt}\delta z = \frac{d}{dt}(\Theta \delta x) &= \dot{\Theta}\delta x + \Theta \frac{d}{dt}\delta x \\ &= \dot{\Theta}\delta x + \Theta \partial_x f \delta x = (\dot{\Theta} + \Theta \partial_x f)\delta x\end{aligned} \tag{24}$$

If Θ is a smooth and bounded transformation, then

$$\frac{d}{dt}\delta z = \underbrace{(\dot{\Theta}\Theta^{-1} + \Theta \frac{\partial f}{\partial x}\Theta^{-1})}_{F}\delta z \tag{25}$$

where F is a generalized Jacobian. We can see that if F is uniformly negative definite, i.e., $F \leq -\alpha I$, then $\dot{\delta z} \leq -\alpha \delta z$, so $\delta z \to 0$ exponentially. Hence, $\delta x \to 0$ exponentially, i.e., $||x_2(t) - x_1(t)|| \to 0$ exponentially with rate α.

Proposition 2 ([8, 2]) *Consider the system $\dot{x} = f(t,x)$. If $\exists \Theta$ with $\Theta^T\Theta \geq \eta I$ for $\eta > 0$, such that the generalized Jacobian F in Eq. 25 is uniformly negative definite, i.e., $F \leq -\alpha I$ for $\alpha > 0$, then the system is contracting and any two trajectories converge to each other exponentially.*

We note that the condition $\Theta^T\Theta \geq \eta I$ ensures invertability of Θ. We also note that as we started from $\delta z = \frac{\partial z}{\partial x}(x,t)\delta x$, this constitutes a differential change of variables that is more general than a simple change of variables $z = z(x,t)$. Indeed, the differential change of variables may not be integrable, hence its generality. Proposition 2 is in fact a necessary and sufficient condition for contraction.
Interestingly,

$$||\delta z||^2 = \delta z^T \delta z = \delta x^T \underbrace{\Theta^T \Theta}_{M(x,t)} \delta x = ||\delta x||_M^2 \tag{26}$$

We see that $M := \Theta^T\Theta$ constitutes a Riemannian metric that redefines the dot product, hence the norm, on $\mathbb{R}^n$. Therefore, if $||\delta z|| \to 0 \implies ||x_2(t) - x_1(t)||_M \to 0$, that is, trajectories tend to each other along geodesics given by the metric M.

An alternative condition to Proposition 2 can also be derived:

$$\frac{d}{dt}\delta z^T \delta z = \frac{d}{dt}\delta x^T M \delta x = \delta x^T(\dot{M} + \partial_x f^T M + M \partial_x f)\delta x$$

Suppose $(\dot{M} + \partial_x f^T M + M\partial_x f) \leq -2\alpha M$, for $\alpha > 0$. Then

$$\frac{d}{dt}\delta x^T M \delta x \leq -2\alpha \delta x^T M \delta x$$

Hence $||\delta x||_M \to 0$ exponentially with rate α. This leads to the following defintion of a contracting system.

Definition 5 ([2, 3, 10, 11]) *Consider the system $\dot{x} = h(x,t)$, a symmetric positive definite metric $M(x) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$, and $\alpha > 0$. If $\forall t \geq 0, \forall x \in \mathbb{R}^n$,*

$$\dot{M} + A^T M + MA \leq -2\alpha M \tag{27}$$

where $A(x,t) = \partial_x h$ and $\dot{M} = \sum_i (\partial_i M) h_i(x)$, then the system is contracting in M with rate α.

We note that, if $M(x) > \beta I$, for $\beta > 0$, then $||x_1(t) - x_2(t)|| \leq \frac{1}{\sqrt{\beta}}||x_1(0) - x_2(0)||_M e^{-\alpha t}$.

4.3 More advanced notions

We here briefly mention a few interesting aspects of contraction theory. We follow [11]. Consider the control-affine nonlinear system $\dot{x} = f(x,t) + B(x,t)u$. We have

$\dot{\delta x} = A(x,u,t)\delta x + B(x,t)\delta u$, where $A = \partial_x f + \sum_i \partial_x b_i u_i$. If there exists a uniformly bounded metric $M(x,t)$, i.e., $\alpha_1 I \leq M(x,t) \leq \alpha_2 I$ such that $\forall \delta x \neq 0$,

$$\delta x^T M B = 0 \implies \delta x^T(\dot{M} + A^T M + MA + 2\alpha M)\delta x < 0 \quad (28)$$

then the system is exponentially stabilizable by continuous feedback with rate α. Such a metric M is called a Control Contraction Metric, CCM. Condition (28) means that the system is naturally contracting in directions orthogonal to the span of the actuated directions. If in (28) A is replaced by $\partial_x f$ and in addition $\forall i, \partial_{b_i} M + \partial_x b_i^T M + M\partial_x b_i = 0$, then a differential stabilizing controller of the form $\delta u = K(x,t)\delta x$ exists. The second condition states that the columns of B form Killing fields for the metric M, i.e., the span of the columns of B preserve distances along geodesics given by M, so that no actuation can expand $||\delta x||$.
To find an appropriate metric M, the contraction along directions orthogonal to the span of actuation and the Killing fields condition can be formulated as convex conditions:

$$-\partial_t W - \partial_f W + \partial_x f W + W \partial_x f^T - \rho B B^T + 2\alpha W < 0$$
$$\partial_{b_i} W - \partial_x b_i W - W \partial_x b_i^T = 0$$

where $W(x,t) = M(x,t)^{-1}$ and $\rho(x,t)$ is a scalar. This leads to a differential feedback gain $K = -\frac{1}{2}\rho B^T W^{-1}$. We refer the reader to [11, 10, 12] for further reading on the theory and application of CCMs.

5 Link between Contraction and Geodesic Convexity

Given the notions introduced above about geodesic convexity and dynamical systems contraction, we can finally make the connection between the two fields. The analysis has been derived in [3].
Consider again the contraction inequality condition (27). Let us re-arrange it into the following form

$$-\frac{1}{2}(\dot{M} + \partial_x h^T M + M\partial_x h) \geq \alpha M$$

Now consider a function f whose Riemannian Hessian is $H = -\frac{1}{2}(\dot{M} + \partial_x h^T M + M\partial_x h)$. Then we would have $H \geq \alpha M$, which is the definition of an α-strongly g-convex function.
It turns out that this is exactly the case for a function f whose natural gradient dynamics $\dot{x} = h(x,t) = -M^{-1}\partial_x f$ are contracting in the metric M, which is the subject of the following theorem.

Theorem 2 ([3]) *Consider a function $f(x,t) \in C^2(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$, a symmetric positive definite metric $M(x) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$, and the natural gradient dynamics*

$$\dot{x} = h(x,t) = -M^{-1}\partial_x f \quad (29)$$

Then, f is α-strongly g-convex in M for each t if and only if (29) is contracting in M with rate α. In this case, the Riemannian Hessian verifies:

$$H = -\frac{1}{2}(\dot{M} + \partial_x h^T M + M\partial_x h) \quad (30)$$

5.1 Some implications

The fundamental connection between geodesic convexity of functions and the contraction of their natural gradient dynamics suggests the possibility to apply contraction analysis tools (most of which were not introduced here because of the limited scope) to geodesically convex optimization. One important application we can readily make is the optimization of non-convex functions. Geodesic convexity implies that there exists a class of functions, sometimes ill-behaved and non-convex, which can be made geodesically convex under a suitable metric M and hence recover favorable certificates, such as the guarantee of global minima. On the other hand, the contraction part suggests a method for finding the global minimum of these functions: perform a natural gradient descent, as opposed to regular gradient descent.
Since a fixed step-size gradient descend algorithm is equivalent to a forward Euler integration scheme for the gradient descent dynamical system, we can analyze the convergence from this perspective. In particular, as the natural gradient dynamics, $\dot{x} = h(x,t) = -M^{-1}\partial_x f$, are contracting in M, and the system is autonomous, then it converges to a stable equilibrium point. Since M is invertible, this point is the unique solution to $\partial_x f = 0$.
Now, even better than a forward Euler integration scheme, one can use a 4th order Runge Kutta dynamical system integration for the natural gradient dynamics. We apply the above discussion on the Rosenbrock function, which is an ill-behaved non-convex function with a single global minimum and is often used to show how good an optimization algorithm is. The function is defined as

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 \quad (31)$$

As shown in Fig. 4, its unique global minimum [1,1] is located in a deep and narrow valley that many algorithms, including descent algorithms, may have a hard time finding it and may get stuck oscillating between the "banks" of the valley.

We compare the performance of the gradient descent with a Runge Kutta integration with that of the natural gradient descent.

- As can be seen in Fig. 5, both descent methods follow different paths to the function's minimum. While the regular gradient descent follows directions of steepest descent on a flat manifold $\mathbb{R}^n$, the natural gradient follows geodesic steepest descent directions on the no-longer-flat $\mathbb{R}^n$ (manifold $(\mathbb{R}^n, M)$).

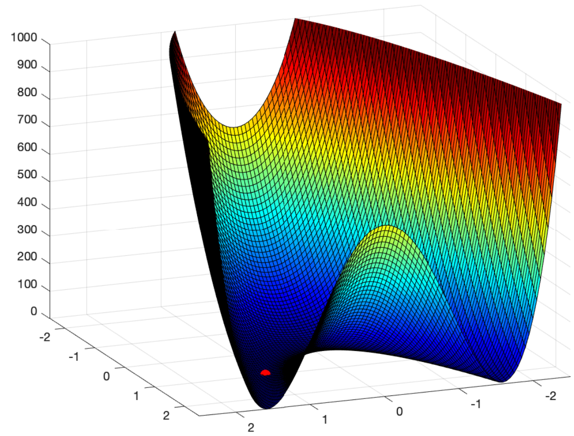

Figure 4: Rosenbrock function with its global minimum [1,1].

- The regular gradient dynamics are extremely sensitive to the step size whereby it no longer converges as soon as the step size $h > 0.002$, whereas the natural gradient dynamics had no trouble converging for step sizes larger than 0.05. Concretely, this can have a big impact on applications that require real-time computations, such as the control of legged robots.

- As seen in Fig. 6, the convergence of the natural gradient descent to the global minimum was 2-3 orders of magnitude faster in terms of number of steps than the regular gradient descent, which reinforces the previous point about applications where online performance is critical.

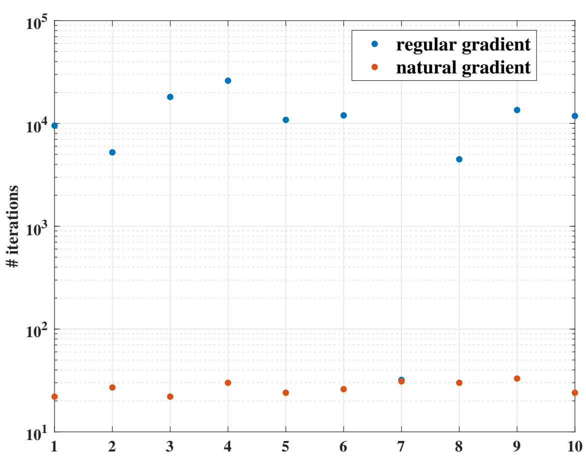

Figure 6: Number of iterations to reach an accuracy of 5e-3 for the regular and the natural gradient descent.

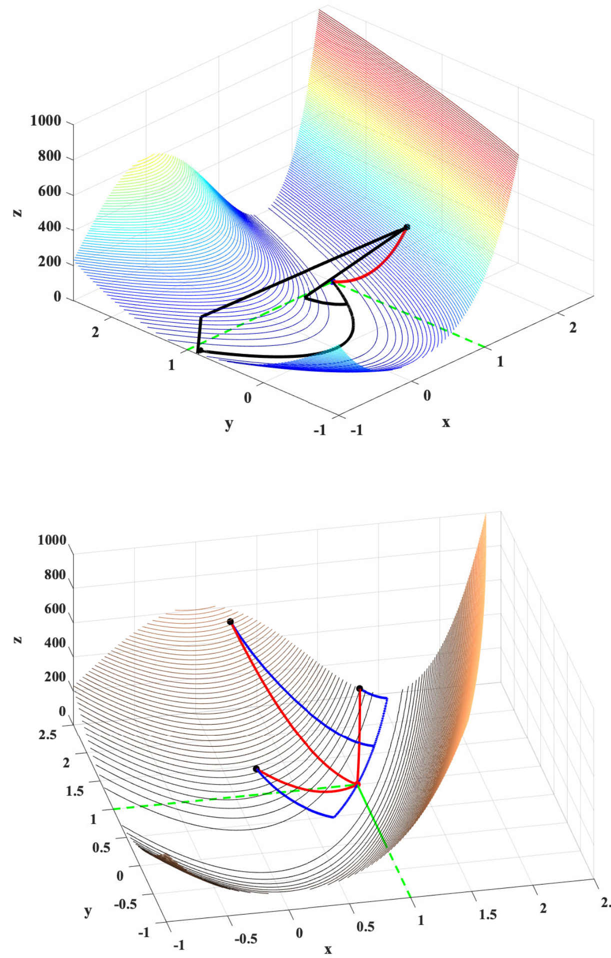

Figure 5: Convergence paths of the regular (black or blue) and the natural gradient descent (red), using (Top) forward Euler where the longest black path is for a step size of 0.002 and the shorter one is for a step size of 0.001, and (Bottom) 4th order, adaptive step, Runge Kutta integration for several initial conditions.

6 Conclusion

Starting from scratch, we have in this project strived to understand the link between two fields, geodesic convexity and contraction analysis. To do so, we sought to study and understand both fields separately, then explore the connection. In brief, there is an interesting link between an extension to the familiar notion of convexity, and a powerful nonlinear systems analysis theory which warrants further investigation. In particular, methods to compute appropriate metrics in this context are still elusive.

References

[1] Constantin Udriste. *Convex functions and optimization methods on Riemannian manifolds*, volume 297. Springer Science & Business Media, 1994.

[2] Winfried Lohmiller and Jean-Jacques E Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998.

[3] Patrick M Wensing and Jean-Jacques E Slotine. Contraction analysis of geodesically convex optimization. *arXiv preprint arXiv:1806.06655*, 2018.

[4] Frederic p schuller's lectures on general relativity.

[5] Youtube lectures on tensor calculus.

[6] Nisheeth K. Vishnoi. Geodesic convex optimization: Differentiation on manifolds, geodesics, and convexity. 2018.

[7] Pavel Pokorny. Geodesic revisited. 2012.

[8] Slotine lectures on nonlinear systems. `http://web.mit.edu/nsl/www/videos/lectures.html`.

[9] Jean-Jacques E Slotine, Weiping Li, et al. *Applied nonlinear control*, volume 199. Prentice hall Englewood Cliffs, NJ, 1991.

[10] Ian R Manchester, Justin Z Tang, and Jean-Jacques E Slotine. Unifying robot trajectory tracking with control contraction metrics. In *Robotics Research*, pages 403–418. Springer, 2018.

[11] Ian R Manchester and Jean-Jacques E Slotine. Control contraction metrics: Convex and intrinsic criteria for nonlinear feedback design. *IEEE Transactions on Automatic Control*, 62(6):3046–3053, 2017.

[12] Sumeet Singh, Anirudha Majumdar, Jean-Jacques Slotine, and Marco Pavone. Robust online motion planning via contraction theory and convex optimization. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5883–5890. IEEE, 2017.