Abstract—Optimization is central to a wide array of problems across machine learning, estimation, automatic control, design, and many other areas. Convex optimization represents a subset of these instances where problems admit global solutions, which can often be computed using off-the-shelf solvers. Recently, increased attention has focused on geodesic convexity, or g-convexity. By endowing \( \mathbb{R}^n \) with a Riemannian metric, non-convex problems can be transformed into convex ones over the induced Riemannian manifold. The main contribution of this paper is to provide a bridge between g-convexity and contraction analysis of nonlinear dynamical systems. Specifically, we show the equivalence between geodesic convexity and contraction of natural gradient flows. Given this result, existing tools from analysis and synthesis of nonlinear contracting systems can be considered to both discover and efficiently optimize g-convex functions. In addition, the contraction perspective allows one to easily apply results to non-autonomous systems, and to recursively build large optimization structures out of simpler elements.

I. INTRODUCTION

This paper considers geodesically-convex optimization through the lens of nonlinear contraction analysis. Many problems in learning, estimation, automatic control and other areas can be phrased in terms of optimization problems posed over Riemannian manifolds \([1], [2]\). Transitioning these problems away from a conventional Euclidean formulation has multiple advantages. Nonlinear constraints that may define the manifold can be eliminated by considering unconstrained formulations on the manifold directly. Second, functions that are not convex in a Euclidean sense may be geodesically convex along shortest paths on the manifold \([3]\).

This second advantage presents opportunity to reformulate non-convex optimization problems over \( \mathbb{R}^n \) as convex ones over a Riemannian manifold. This reformulation is accomplished not through any explicit coordinate change, but simply by endowing \( \mathbb{R}^n \) with a suitable Riemannian metric. Principled search over metrics can be accomplished computationally \([4]\) and represents additional generality beyond considering nonlinear changes of coordinates. In effect, endowing \( \mathbb{R}^n \) with a Riemannian metric may be viewed as following from a differential change of coordinates \([5]\).

While an explicit coordinate change can be very effective if it fundamentally derives from the structure or the physics of the system, a differential coordinate change is not necessarily integrable and thus is considerably more general. A similar approach is at the heart of nonlinear contraction analysis. Contraction theory \([5]\) allows stability considerations of a nonlinear non-autonomous system on \( \mathbb{R}^n \) to be transformed into those of a linear time-varying one along the system’s flow. The existence of a Riemannian metric that contracts virtual displacements (i.e., elements in the tangent space) is necessary and sufficient for exponential convergence of any pair of trajectories. Contraction naturally yields methods for constructing stable systems of systems, including synchronization phenomena and consensus as well as other key building-blocks that allow the construction of large contracting systems out of simpler elements.

The main contribution of this paper is to provide a precise link between the geodesic convexity of functions on \( \mathbb{R}^n \) and the contraction of their gradient flows. Section \( \text{III} \) introduces key definitions and our main theoretical result. Specifically, we analyze continuous time gradient descent on a manifold, which takes the form of the natural gradient \([6]\) in coordinates. Analysis in continuous time optimization is limited, in part, by the fact that dynamics can be arbitrarily sped up to achieve any convergence rate without regards for how it may affect a discrete implementation. However, a continuous perspective has yielded insight on important phenomena such as in the analysis \([7]\) and extensions \([8]\) of Nesterov’s accelerated gradient descent method \([9]\). It has also enabled analysis of primal-dual algorithms \([10]\), where an absolute time reference is obtained by introducing additional fast dynamics or delays using a singular perturbation framework.

Following the main result for g-convexity in Section \( \text{II} \) Section \( \text{III} \) presents an extension for the analysis of primal-dual type dynamics that appear in mixed convex/concave saddle systems. Section \( \text{IV} \) then discusses how these insights can be used to design larger optimization structures out of simpler ones by leveraging combination properties of contracting systems. Section \( \text{V} \) provides an outlook on potential future advances that may stem from these connections.

II. DEFINITIONS AND MAIN RESULT

We first recall basic definitions and facts on convex and g-convex functions, and on contraction analysis of nonlinear dynamical systems. These definitions provide the context for our main result, Theorem \( \text{I} \) that establishes equivalence between g-convexity and contraction of gradient flows.

A. Convex Optimization

**Definition 1** (Strong Convexity). A function \( f \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}) \) is \( \alpha \)-strongly convex with \( \alpha > 0 \) if its Hessian matrix \( \partial_{xx} f \) satisfies the matrix inequality

\[
\partial_{xx} f \succeq \alpha I \quad \forall x \in \mathbb{R}^n
\]

As its name suggests, a function that is strongly convex is convex in the usual sense, while the converse is not always true. From a dynamic systems perspective, strong convexity provides exponential convergence of gradient flows:
Proposition 1. If a function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is $\alpha$-strongly convex, then its gradient system
\begin{equation}
\dot{x} = -\partial_x f
\end{equation}
converges to the global minimum exponentially with rate $\alpha$. 

B. g-Convex Optimization

Geodesic convexity generalizes conventional notions of convexity to the case where the domain of a function is viewed as a Riemannian manifold. A special case occurs in geometric programming [11] where non-convex problems that can be transformed into convex ones by a change of variables. Geodesically-convex optimization generalizes these insights to a broader class of problems [2]. However, beyond special cases (see, e.g., [12]), generative procedures remain lacking to formulate g-convex optimization problems or recognize g-convexity.

Geodesic-convexity of a function over $\mathbb{R}^n$ is not an intrinsic property of the function itself, but of a function paired with a suitable positive definite metric $M(x) : \mathbb{R}^n \to \mathbb{R}^{n \times n}$. The pair $(\mathbb{R}^n, M)$ is a Riemannian manifold.

Definition 2 (g-Strong Convexity [3]). A function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is said to be geodesically $\alpha$-strongly convex (with $\alpha > 0$) in a symmetric positive definite metric $M(x)$ if its Riemannian Hessian matrix $H(x)$ satisfies:
\begin{equation}
H(x) \succeq \alpha M(x) \quad \forall x \in \mathbb{R}^n
\end{equation}
The elements of the Riemannian Hessian are given as
\begin{equation}
H_{ij} = \partial_{ij} f - \Gamma_{ij}^k \partial_k f
\end{equation}
where $\partial_{ij} f = \frac{\partial^2 f}{\partial x_i \partial x_j}$ provide the elements of the conventional (Euclidean) Hessian and $\Gamma_{ij}^k$ denotes the Christoffel symbols of the second kind
\begin{equation}
\Gamma_{ij}^k = \frac{1}{2} \sum_{k=1}^n (M^{mk} (\partial_j M_{ik} + \partial_i M_{jk} - \partial_k M_{ij}))
\end{equation}
with $M^{ij} = (M^{-1})_{ij}$. The function $f$ is g-convex when [2] holds with $\alpha = 0$.

Note that when $M(x)$ is the identity metric, geodesic $\alpha$-convexity coincides with the definition of $\alpha$-strong convexity in Definition 1. Toward providing a parallel to Proposition 1 within the Riemannian context, we consider stability analysis in the Riemannian context through the lens of contraction theory.

Definition 3 (Contraction Metric [5]). A system $\dot{x} = h(x, t)$ is said to be contracting at rate $\alpha > 0$ with respect to a symmetric positive definite metric $M(x, t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n \times n}$, if for all $t \geq 0$ and all $x \in \mathbb{R}^n$,
\begin{equation}
\dot{M} + A^T M + M A \preceq -2\alpha M
\end{equation}
where $A(x, t) = \partial_x h$ is the system Jacobian and $\dot{M} = \partial_t M + \sum_{i} (\partial_i M) h_i(x)$. The system is said to be semi-contracting with respect to $M(x, t)$ when 3 holds with $\alpha = 0$.

As a key property, given an $\alpha$-contracting system and an arbitrary pair of initial conditions $x_1(0)$ and $x_2(0)$, the solutions $x_1(t)$ and $x_2(t)$ converge to one another exponentially
\begin{equation}
d_{M(t)}(x_1(t), x_2(t)) \leq e^{-\alpha t} d_{M(0)}(x_1(0), x_2(0))
\end{equation}
where $d_{M(t)}(\cdot, \cdot)$ denotes the geodesic distance on the Riemannian manifold $M(t) = (\mathbb{R}^n, M(x, t))$. Property 3 can be shown [5] by analyzing differential displacements $\delta x$ at fixed time between trajectories, which verify
\begin{equation}
\frac{d}{dt}(\delta x^T M \delta x) \leq -2\alpha \|\delta x\|^2
\end{equation}
Furthermore, if the metric satisfies $M(x, t) \geq \beta I$ uniformly in $(x, t)$ for some constant $\beta > 0$, then any two solutions of an $\alpha$-contracting system verify
\begin{equation}
\|x_1(t) - x_2(t)\| \leq \frac{1}{\sqrt{\beta}} e^{-\alpha t} (x_1(0) - x_2(0))
\end{equation}

Example 1. Consider an $\alpha$-strongly convex function $f$ and its associated gradient descent system [1]. Since $f$ is strongly convex, it has a unique global minimum $x^\ast$, which is a stationary point of $f$. It can be verified that the gradient descent dynamics of $f$ are contracting in an identity metric $M = I$ with rate $\alpha$. Since geodesic distances are just Euclidean distances in this metric, 3 immediately implies that
\begin{equation}
\forall t \geq 0, \|x(t) - x^\ast\| \leq e^{-\alpha t} \|x(0) - x^\ast\|
\end{equation}
thus proving Proposition 1.

The following theorem provides a critical link between g-convexity and contraction analysis of nonlinear systems, mirroring Proposition 1 within the Riemannian context.

Theorem 1. Consider a function $f(x) \in C^2(\mathbb{R}^n, \mathbb{R})$, a symmetric positive definite metric $M(x) : \mathbb{R}^n \to \mathbb{R}^{n \times n}$, and the natural gradient system [6]
\begin{equation}
\dot{x} = h(x) = -M(x)^{-1} \partial_x f
\end{equation}
Then, $f$ is $\alpha$-strongly g-convex in the metric $M$ if and only if 3 is contracting with rate $\alpha$ in the metric $M$. More specifically, the Riemannian Hessian verifies
\begin{equation}
H = -\frac{1}{2} \left( M (\partial_x h) + (\partial_x h)^T M + \dot{M} \right)
\end{equation}
Proof. See Appendix 1. □

Remark 1. While the above theorem applies to $\alpha$-strong convexity, the link between the Riemannian Hessian and the contraction condition 3 also provides immediate equivalence between g-convexity of a function and semi-contracting of its natural gradient dynamics.

Remark 2. The above result provides an alternate way to compute the geodesic Hessian $H(x)$, and, as expected, leaves it invariant when the metric $M(x)$ is scaled by a strictly positive constant. Because of the structure of the natural gradient dynamics, scaling $M(x)$ is akin to scaling
time and implies inversely scaling the contraction rate $\alpha$, consistently with $\ref{eq:alpha}$. Let us illustrate this result using the classical nonconvex Rosenbrock function:

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 \quad \text{(7)}$$

This function has a unique global optimum at $x^* = [1, 1]^T$, which is located along a long, shallow, parabolic-shaped valley. As a result, it exhibits poor scaling and is frequently used as a test problem for optimization.

**Example 2.** Consider the Rosenbrock function $\ref{eq:rosenbrock}$ and the metric

$$\mathbf{M}(x) = \begin{bmatrix} 400x_1^2 + 1 & -200x_1 \\ -200x_1 & 100 \end{bmatrix}$$

The metric $\mathbf{M}(x)$ satisfies $\text{tr}(\mathbf{M}(x)) = 400x_1^2 + 101 > 0$ and $\det(\mathbf{M}(x)) = 100 > 0$, and thus $\mathbf{M}(x) \succ 0$. Note that $\mathbf{M}(x)$ is not simply the Hessian of $f(x)$. The natural gradient dynamics follows

$$\dot{x} = \mathbf{h}(x) = -\mathbf{M}(x)^{-1}(\partial_x f) \quad = -2 \begin{bmatrix} x_1 - 1 \\ x_1^2 - 2x_1 + x_2 \end{bmatrix}$$

It can be verified algebraically that

$$\mathbf{M}(\partial_x \mathbf{h}) + (\partial_x \mathbf{h})^\top \mathbf{M} + \mathbf{M} = -4\mathbf{M}$$

which shows that the system is contracting with rate $\alpha = 2$. This implies that the natural gradient dynamics satisfies

$$d_{\mathbf{M}}(x(t), x^*) \leq e^{-2t} d_{\mathbf{M}}(x(0), x^*)$$

where $x^* = [1, 1]^T$. Equivalently, the Rosenbrock function is geodesically $\alpha$-strongly convex with $\alpha = 2$.

The Rosenbrock metric $\mathbf{M}(x)$ can be viewed as following from a differential change of variables

$$\delta z = \Theta(x) \delta x = \begin{bmatrix} 20x_1 & -10 \\ 1 & 0 \end{bmatrix} \delta x$$

where $\mathbf{M} = \Theta^\top \Theta$ yields $\delta x^\top \mathbf{M} \delta x = \delta x^\top \delta z$. This differential change of variables is integrable, so that $g$-convexity of the Rosenbrock can be shown using the explicit nonlinear coordinate change $z_1 = 10x_1^2 - 10x_2$ and $z_2 = x_1 - 1$ that provides $f = z_1^2 + z_2^2$. Although in this case the coordinate transform is integrable, the general freedom to consider differential changes of coordinates provides additional flexibility and generality to both contraction analysis and $g$-convexity.

**Example 3.** Geodesically-convex optimization can also be used to carry out manifold-constrained optimization in an unconstrained fashion via recasting problems over a Riemannian manifold $\ref{eq:riemannian}$.

Consider for instance optimization over the set $\mathbb{S}_n^+$ of $n \times n$ positive definite matrices, and specifically the problem of finding the Karcher mean of $n$ matrices $\mathbf{A}_i \in \mathbb{S}_n^+$ $\ref{eq:karcher}$, which minimizes the objective function

$$f(\mathbf{X}) = \frac{1}{2} \sum_{i=1}^n \|\log(\mathbf{X}^{-1/2} \mathbf{A}_i \mathbf{X}^{-1/2})\|_F^2$$

where $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^\top \mathbf{A})}$ denotes the Frobenius norm of a matrix $\mathbf{A}$. The function $f(\mathbf{X})$ is $m$-strongly convex $\ref{eq:strongly}$ on $\mathbb{S}_n^+$ in the metric which measures square differential displacements as

$$\text{tr}(\delta \mathbf{X} \mathbf{X}^{-1} \delta \mathbf{X} \mathbf{X}^{-1}) \quad \text{(8)}$$

This metric coincides with the Hessian of the log barrier $-\log \text{det}(\mathbf{X})$ $\ref{eq:log_barrier}$. The gradient of $f(\mathbf{X})$ can be written

$$\partial_x f = \sum_{i=1}^m \mathbf{X}^{-1/2} \log(\mathbf{X}^{1/2} \mathbf{A}_i^{-1} \mathbf{X}^{1/2}) \mathbf{X}^{-1/2}$$

and accordingly the natural gradient can be shown to satisfy

$$\sum_{i=1}^m \mathbf{X}^{1/2} \log(\mathbf{X}^{1/2} \mathbf{A}_i^{-1} \mathbf{X}^{1/2}) \mathbf{X}^{1/2} = \mathbf{X} (\partial_x f) \mathbf{X}$$

From Theorem $\ref{eq:strongly}$ any trajectory with arbitrary initial condition $\mathbf{X}(0) \in \mathbb{S}_n^+$ will remain within $\mathbb{S}_n^+$ under the natural gradient descent dynamics

$$\dot{\mathbf{X}} = -\sum_{i=1}^m \mathbf{X}^{1/2} \log(\mathbf{X}^{1/2} \mathbf{A}_i^{-1} \mathbf{X}^{1/2}) \mathbf{X}^{1/2}$$

since (intuitively) the Riemannian metric $\ref{eq:rosenbrock}$ makes positive semi-definite matrices an infinite distance away from any positive definite one, and contraction of the natural gradient dynamics ensures that geodesic distances decrease exponentially.

Note that in the autonomous case, equations governing the differential displacement at fixed time $\delta x$ satisfy

$$\frac{d}{dt} \delta x = \partial_x \mathbf{h}(x) \delta x \quad \text{(9)}$$

which has a similar structure to the time evolution of $\mathbf{h}(x)$

$$\frac{d}{dt} \mathbf{h}(x) = \partial_x \mathbf{h}(x) \mathbf{h}(x) \quad \text{(10)}$$

Thus, for natural gradient descent of an $\alpha$-strong g-convex function $f(x)$, the same algebra leading to $\ref{eq:contraction}$ also gives

$$\frac{d}{dt} (\mathbf{h}^\top \mathbf{M} \mathbf{h}) \leq -2\alpha (\mathbf{h}^\top \mathbf{M} \mathbf{h})$$

so that

$$\mathbf{V}(x) = \mathbf{h}(x)^\top \mathbf{M}(x) \mathbf{h}(x) = (\partial_x f)^\top \mathbf{M}(x)^{-1} \partial_x f$$

can be viewed as an exponentially converging Lyapunov function, with global minimum $V = 0$ at the unique minimum of $f(x)$. Of course, $\ref{eq:contract}$ remains valid for non-autonomous systems as well, while $\ref{eq:autonomous}$ does not.

**C. Non-autonomous systems**

In our optimization context, the fact that contraction analysis is directly applicable to non-autonomous systems can be exploited in a variety of ways. While the most immediate is that it can address optimization problems set in time-varying environments, as e.g. in $\ref{eq:autonomous}$, a key aspect is that it allows combination properties of contracting systems (such as parallel and feedback combinations, as well as hierarchies or cascades) to be exploited to address more
elaborate optimization problems or architectures. Also, as we shall see later, it makes possible the construction of virtual systems [14] to potentially extend results beyond strict natural-gradient descent or strictly contracting systems.

III. GEODESIC PRIMAL-DUAL OPTIMIZATION

The Primal-Dual algorithm is widely used in optimization to determine saddle points. Saddle points generally occur when a system is simultaneously minimizing a function over one set of its variables and maximizing it over another variable set. Saddle points appear in economics in the context of market equilibrium [15], and also appear naturally in constrained optimization [13], where Lagrange parameters play the role of dual variables. When a function is strictly convex in a subset of its variables, and strictly concave in the remaining, gradient descent/ascent dynamics converge to a unique saddle equilibrium [15], [16]. Within the context of constrained optimization, these dynamics are known as the primal-dual dynamics. Such dynamics play an important role e.g. in machine learning, for instance in the adversarial in constrained optimization [13], where Lagrange parameters is replaced by $g$-convexity, thus broadening dual optimization, where convexity in terms of primal and dual variables is replaced by $g$-convexity, thus broadening the above results to state-dependent metrics.

Consider a scalar cost function $\mathcal{L}(x, \lambda, t)$, possibly time-dependent, with $\mathcal{L}$ $g$-strongly convex over $x$ and $g$-strongly concave over $\lambda$ in metrics $M_x(x)$ and $M_\lambda(\lambda)$ respectively. Consider the geodesic primal-dual dynamics, which we define as

$$
M_x(x) \dot{x} = -\partial_x \mathcal{L} \quad (11a)
$$

$$
M_\lambda(\lambda) \dot{\lambda} = -\partial_\lambda \mathcal{L} \quad (11b)
$$

Using the metrics $M_x(x)$ and $M_\lambda(\lambda)$ extends the standard case [10], where they would be replaced by constant, symmetric positive definite matrices.

Theorem 2. The geodesic primal-dual dynamics (11) is globally contracting, with metric

$$
M(x, \lambda) = \begin{bmatrix} M_x(x) & 0 \\ 0 & M_\lambda(\lambda) \end{bmatrix} \quad (12)
$$

Proof. Letting $z = [x^\top, \lambda^\top]^\top$ and $\dot{z} = f(z)$ denote the overall system dynamics, the system Jacobian can be written

$$
A(x, \lambda, t) = \partial_z f = \begin{bmatrix} -\partial_x (M_x^{-1} \partial_x \mathcal{L}) & -M_x^{-1} \partial_x \mathcal{L} \\ M_\lambda^{-1} \partial_\lambda \mathcal{L} & -\partial_\lambda (M_\lambda^{-1} \partial_\lambda \mathcal{L}) \end{bmatrix} \quad (13)
$$

so that, using Theorem 1

$$
M^\top A + A^\top M + M = -2 \begin{bmatrix} H_x & 0 \\ 0 & -H_\lambda \end{bmatrix} < 0
$$

IV. APPLYING CONTRACTION ANALYSIS TOOLS TO GEODESIC CONVEXITY

Theorem 1 immediately implies that existing contraction tools can be directly applied in the context of geodesic optimization. These include, in particular, combination properties and virtual systems.

A. Combination Properties

We now discuss how basic combination properties of contracting systems [5], [22] can be exploited to build on the results above. While these properties derive from simple matrix algebra and in principle could be proven directly from the definition of geodesic convexity, as we will see most rely for their practical relevance on the flexibility afforded by the contraction analysis point of view.

1) Sum of $g$-convex: If two functions $f_1(x)$ and $f_2(x)$ are $g$-convex in the same metric, then their sum $f_1(x) + f_2(x)$ is $g$-convex in the same metric.

Example 4. Consider a function $f_1(x_1, y_1)$ $g$-convex in a block diagonal metric $\text{BlkDiag}(M_{x_1}, M_y(y_1))$ and a function $f_2(x_2, y_2)$ $g$-convex in a block diagonal metric $\text{BlkDiag}(M_{x_2}(x_2), M_y(y_2))$. Then, the function:

$$
f(x_1, x_2, y) = f_1(x_1, y) + f_2(x_2, y)
$$

is $g$-convex in the metric $\text{BlkDiag}(M_{x_1}, M_{x_2}, M_y)$.

2) Skew-Symmetric Feedback Coupling: The system Jacobian (13) for the primal-dual dynamics is a special case of a contraction property that holds more broadly.

Assume that a scalar function $f_1(x_1, x_2)$ is $\alpha_1$-strongly $g$-convex in $x_1$ in a metric $M_1(x_1)$ for each fixed $x_2$, and similarly that a scalar function $f_2(x_1, x_2)$ is $\alpha_2$-strongly $g$-convex in a metric $M_2(x_2)$ for each fixed $x_1$. If $f_1$ and $f_2$ satisfy the scaled skew-symmetry property

$$
\partial_{x_1, x_2} f_1 = -k (\partial_{x_2, x_1} f_2)^\top
$$

where $k$ is some strictly positive constant, then the natural gradient dynamics

$$
\begin{align*}
\dot{x}_1 &= -M_1(x_1)^{-1} \partial_{x_1} f_1(x_1, x_2) \\
\dot{x}_2 &= -M_2(x_2)^{-1} \partial_{x_2} f_2(x_2, x_1)
\end{align*} \quad (14)
$$

is contracting with rate $\min(\alpha_1, \alpha_2)$ in metric $M(x_1, x_2) = \text{BlkDiag}(M_1(x_1), kM_2(x_2))$. Since the overall system is both contracting and autonomous, it tends to a unique equilibrium [5] $(x_i^*, x_j^*)$ which satisfies the Nash-like conditions

$$
\begin{align*}
x_i^* &= \text{argmin}_{x_i} f_1(x_1, x_i^*) \\
x_j^* &= \text{argmin}_{x_j} f_2(x_j^*, x_2)
\end{align*}
$$

This result extends to a game with an arbitrary number of players. Consider $n$ functions $\{f_i(x_1, \ldots, x_n)\}_{i=1}^n$ such that
each \( f_i \) is \( \alpha_i \)-strongly \( g \)-convex over \( x_i \) in a metric \( M_i(x_i) \). If the functions satisfy the skew-symmetry conditions
\[
\partial_{x_j x_i} f_i = -k_{ij} (\partial_{x_j} x_i, f_j) ^{\top}
\]
for each \( j > i \), then the suitable generalizations of (14) result in a coupled system that is contracting with rate \( \min(\alpha_1, \ldots, \alpha_n) \) in the metric \( M = \text{BlkDiag}(M_1, k_2 M_2, \ldots, k_n M_n) \). The overall system converges to a unique Nash-like equilibrium satisfying
\[
x^*_i = \text{argmin}_{x_i} f_1(x_1, x_2^*, \ldots, x_n^*)
\]
with similar results for each other player.

3) Hierarchical Natural Gradient: Consider a function \( f_1(x_1) \) \( \alpha_1 \)-strongly \( g \)-convex in a metric \( M_1(x_1) \), and a function \( f_2(x_1, x_2) \) \( \alpha_2 \)-strongly \( g \)-convex in a metric \( M_2(x_2) \) for each given \( x_1 \), where the \( x_i \) may have different dimensions. Then, the hierarchical natural gradient dynamics
\[
\begin{align*}
\dot{x}_1 &= -M_1(x_1)^{-1} \partial_{x_1} f_1(x_1) \\
\dot{x}_2 &= -M_2(x_2)^{-1} \partial_{x_2} f_2(x_1, x_2)
\end{align*}
\]
is contracting with rate \( \min(\alpha_1, \alpha_2) \) in metric \( M(x_1, x_2) = \text{BlkDiag}(M_1(x_1), M_2(x_2)) \), under the mild assumption that the coupling Jacobian is bounded [5]. Since the overall system is both contracting and autonomous, it tends to a unique equilibrium [5] at rate \( \min(\alpha_1, \alpha_2) \), and thus to the unique solution of
\[
\begin{align*}
\partial_{x_1} f_1(x_1) &= 0 \\
\partial_{x_2} f_2(x_1, x_2) &= 0
\end{align*}
\]
By recursion, this structure can be chained an arbitrary number of times, or applied to any cascade or directed acyclic graph of natural gradient dynamics. Such hierarchical optimization may play a role, for instance, in backpropagation of natural gradients in machine learning, with all descents occurring concurrently rather than in sequence.

B. Virtual Systems

The use of virtual contracting systems [14], [23], [24] allows guaranteed exponential convergence to a unique minimum to be extended to classes of dynamics which are not pure natural gradient.

Consider for instance the natural gradient descent (6) with the function \( f(x) \) \( \alpha \)-strongly \( g \)-convex in metric \( M(x) \), and define the new system
\[
\dot{x} = -p(x, t) M(x)^{-1} \partial_x f(x)
\]
where the smooth scalar function \( p(x, t) \) is uniformly positive definite, \( \exists\ p_{\text{min}} > 0, \forall t \geq 0, \forall x, \ p(x, t) \geq p_{\text{min}} \)

Let us show now that this system tends exponentially to the minimum of \( f(x) \).

Consider the auxiliary, virtual system,
\[
\dot{y} = -M(y)^{-1} \partial_y (p(x, t) f(y))
\]

For this system, \( p(x(t), t) \) is an external, uniformly positive definite function of time, and thus
\[
\partial_y (p(x, t) f(y)) = p(x, t) \partial_y f(y)
\]
so that the contraction of (6) with rate \( \alpha \) implies contraction of (16) with rate \( \alpha p_{\text{min}} \) to the same global optimizer.

Note that since we only assumed that \( p(x, t) \) is uniformly positive definite, in general the actual system (15) is not contracting with respect to the metric \( M(x) \). Also, the approach extends immediately to the primal-dual context of Section III.

V. Conclusions

This paper has outlined a fundamental connection between \( g \)-strongly convex functions and contraction of their natural gradient dynamics. This observation sets the foundation for contraction results to be exploited for the construction and analysis of \( g \)-convex optimization for complex systems. A natural next step for the application of contraction in \( g \)-convex optimization is to design geodesic quorum sensing [25], [26] synchronization algorithms, as well as other consensus mechanisms considering time-delays [27], [28], which may serve as the basis for distributed and large-scale optimization techniques on Riemannian manifolds. Such advances could have clear impact e.g. in the context of machine learning [29], among other applications.

APPENDIX I

PROOF OF THEOREM I

\textbf{Proof.} Recall that \( \alpha \)-strong geodesic convexity of \( f(x) \) in the metric \( M(x) \) is equivalent to the Riemannian Hessian of \( f \), denoted \( H(x) \), satisfying:
\[
H \succeq \alpha M
\]

We now show that this property is exactly the same as contraction of the natural gradient dynamics (6) in the metric \( M(x) \). Specifically, given \( h(x) \) from (6), and defining
\[
Q = M (\partial_x h) + (\partial_x h)^{\top} M + \dot{M}
\]
we show that \( Q = -2H \), thus proving the theorem.
In coordinates, entries of \( H \) are given by
\[
H_{ij} = \partial_{ij} f - \Gamma^k_{ij} (\partial_k f)
\]
where \( \Gamma^k_{ij} \) denotes the Christoffel symbol of the second kind
\[
\Gamma^k_{ij} = \frac{1}{2} M^{mk} (\partial_j M_{ik} + \partial_i M_{jk} - \partial_k M_{ij})
\]
To streamline derivation, Einstein summation notation is used, e.g., implying a sum over \( k \) in the above.
Consider the partials of the natural gradient system,
\[
\partial_j h_k = \partial_j \left[ -M^{kl} (\partial_k f) \right]
\]
\[
= -M^{kl} (\partial_j f) + M^{kl} (\partial_j M_{rs}) M^{st} (\partial_t f)
\]
Using this result
\[
Q_{ij} = M_{ik}(\partial_j h_k) + M_{jk}(\partial_i h_k) - (\delta_{ik}M_{ij})M^{jk}(\partial_i f) \\
= -M_{ik}M^{jk}(\partial_j f) + M_{ik}M^{kr}(\partial_j M_{kr})M^{jk}(\partial_i f) \\
- M_{ik}M^{jk}(\partial_i k) + M_{ik}M^{kr}(\partial_i M_{kr})M^{jk}(\partial_j f) \\
- (\delta_{ik}M_{ij})M^{jk}(\partial_i f)
\]
Noting that \(M_{ik}M^{jk} = \delta_{ij}\), with \(\delta_{ij}\) the Kronecker delta,
\[
Q_{ij} = -20\delta_{ij} f + M^{ik}(\partial_i M_{ks} + \partial_i M_{js} - \delta_{k}M_{ij}) \partial_i f \\
= -2(\partial_{ij} f - \Gamma^{i}_{ij} \partial_i f) \\
= -2H_{ij}
\]

**References**

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. [Online]. Available: [http://www.jhu.edu/group/stable/jct/5mmk]

[2] H. Zhang and S. Sra, “First-order methods for geodesically convex optimization,” in *Conference on Learning Theory*, 2016, pp. 1617–1638.

[3] C. Udriste, *Convex functions and optimization methods on Riemannian manifolds*. Springer Science & Business Media, 1994, vol. 297.

[4] M. A. Aylward, P. A. Parrilo, and J.-J. E. Slotine, “Stability and robustness analysis of nonlinear systems via contraction metrics and sos programming,” *Automatica*, vol. 44, no. 8, pp. 2163–2170, 8 2008. [Online]. Available: [http://www.sciencedirect.com/science/article/pii/S0005109808000563]

[5] W. Lohmiller and J.-J. E. Slotine, “On contraction analysis for nonlinear systems,” *Automatica*, vol. 34, no. 6, pp. 683–696, 6 1998.

[6] S.-I. Amari, “Natural gradient works efficiently in learning,” *Neural Comput.*, vol. 10, no. 2, pp. 251–276, Feb. 1998. [Online]. Available: [http://dx.doi.org/10.1162/089976698300017746]

[7] W. Su, S. Boyd, and E. Candès, “A differential equation for modeling nesterov’s accelerated gradient method: Theory and insights,” in *Advances in Neural Information Processing Systems*, 2014, pp. 2510–2518.

[8] A. Wibisono, A. C. Wilson, and M. I. Jordan, “A variational perspective on accelerated methods in optimization,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 47, pp. E7351–E7358, 2016.

[9] Y. Nesterov, *Introductory lectures on convex programming – A Basic course*. Springer, 1998.

[10] H. D. Nguyen, T. L. Vu, K. Turitsyn, and J.-J. Slotine, “Contraction and robustness of continuous time primal-dual dynamics,” *IEEE Control Systems Letters*, vol. 2, no. 4, 2018.

[11] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, “A tutorial on geometric programming,” *Optimization and Engineering*, vol. 8, no. 1, p. 67, Apr. 2007. [Online]. Available: [https://doi.org/10.1007/s11081-007-9001-7]

[12] S. Sra and R. Hosseini, “Conic geometric optimization on the manifold of positive definite matrices,” *SIAM Journal on Optimization*, vol. 25, no. 1, pp. 713–739, 2015. [Online]. Available: [https://doi.org/10.1137/140978168]

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[14] W. Wang and J.-J. E. Slotine, “On partial contraction analysis for coupled nonlinear oscillators,” *Biological cybernetics*, vol. 92, no. 1, pp. 38–53, 2005.

[15] K.J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Nonlinear Programming*. Stanford University Press, Stanford, CA, 1958.

[16] D. Feijer and F. Paganini, “Stability of primal–dual gradient dynamics and applications to network optimization,” *Automatica*, vol. 46, no. 12, pp. 1974 – 1981, 2010.

[17] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards deep learning models resistant to adversarial attacks,” *arXiv preprint arXiv:1706.06083*, 2017.

[18] K. C. Kosaraju, S. Mohan, and R. Pasumarthy, “On the primal–dual dynamics of support vector machines,” *arXiv preprint arXiv:1805.00699*, 2018.

[19] W. S. Cho and M. Wang, “Deep primal-dual reinforcement learning: Accelerating actor-critic using bellman duality,” *CoRR*, vol. abs/1712.02467, 2017. [Online]. Available: [http://arxiv.org/abs/1712.02467]

[20] R. Ortega, D. Jeltensena, and J. M. Scherpen, “Power shaping: A new paradigm for stabilization of nonlinear lc circuits,” *IEEE Transactions on Automatic Control*, vol. 48, no. 10, pp. 1762–1767, 2003.

[21] K. Cavanagh, J. A. Belk, and K. Turitsyn, “Transient stability guarantees for ad hoc dc microgrids,” *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 139–144, 2018.

[22] J.-J. Slotine, “Modular stability tools for distributed computation and control,” *International Journal of Adaptive Control and Signal Processing*, vol. 17, no. 6, pp. 397–416, 2003.

[23] J. Jouffroy and J.-J. Slotine, “Methodological remarks on contraction theory,” in *Decision and Control, 2004. CDC 43rd IEEE Conference on*, vol. 3, Dec 2004, pp. 2537–2543 Vol.3.

[24] I. R. Manchester and J. J. E. Slotine, “Control contraction metrics: Convex and intrinsic criteria for nonlinear feedback design,” *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 3046–3053, June 2017.

[25] N. Tabareau, J.-J. Slotine, and Q.-C. Pham, “How synchronization protects from noise,” *PLoS computational biology*, vol. 6, no. 1, p. e1000637, 2010.

[26] G. Russo and J. J. E. Slotine, “Global convergence of quorum-sensing networks,” *Physical Review E*, vol. 82, no. 4, p. 041919, 2010.

[27] W. Wang and J. J. E. Slotine, “Contraction analysis of time-delayed communications and group cooperation,” *IEEE Transactions on Automatic Control*, vol. 51, no. 4, pp. 712–717, April 2006.

[28] P. M. Wensing and J. J. E. Slotine, “Cooperative Adaptive Control for Cloud-Based Robotics,” *Proceedings of the IEEE International Conference on Robotics and Automation*, 2018.

[29] S. Zhang, A. Choromanska, and Y. LeCun, “Deep learning with elastic averaging sgd,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS’15. Cambridge, MA, USA: MIT Press, 2015, pp. 685–693. [Online]. Available: [http://dl.acm.org/citation.cfm?id=2969239.2969316]