Geodesic Convexity

Arpon Basu

Indian Institute of Technology, Bombay

April 19, 2024

Need for Geodesic Convexity

1. Convex optimization (over $\mathbb{R}^n$) very powerful, but not applicable in many cases
2. Special class of non-convex functions becomes convex once one imposes a different manifold structure on the underlying domain. Indeed, consider the function $f : \mathbb{R}^n_{>0} \mapsto \mathbb{R}$, where:

$$f(x_1, \ldots, x_n) := \ln(x_1^2 x_2^2 \cdots x_n^2 + x_n^{2n}) - \sum_{i=1}^{n} \ln(x_i)$$

   f is clearly not convex. However, under the manifold structure induced by the Hessian of the function $-\sum_{i=1}^{n} \ln(x_i)$, the function is *geodesically convex*
3. Many more non-convex functions arising in very natural contexts were found to be geodesically convex when the 'correct' Riemannian metric tensor was imposed on the manifold (instead of the usual Euclidean metric tensor).

We first define a (pseudo)-Riemannian manifold.

Definition

A (pseudo)-Riemannian manifold is a smooth n-dimensional manifold M equipped with a smooth function $g : M \mapsto \mathcal{S}_n$, where $\mathcal{S}_n$ is the manifold of $n \times n$ symmetric invertible matrices.

The map g is sometimes also called the *metric tensor*: Indeed, for any $p \in M$, we have an inner product on T_pM, given by $\langle u, v \rangle_g := u^\top G(p) v$, where T_pM is identified with $\mathbb{R}^n$. Clearly, once we have an inner product, we can also define a norm, and hence a metric. However, the reader is asked to note that the metric will be non-negative only if the image of g lies in the space of positive-definite matrices (in which case we call M a Riemannian manifold).

Christoffel Symbols

The definition of a metric tensor also allows us to define the Christoffel symbols:

Definition (Christoffel Symbols)

Fix a point p. $G(p)$ is a $n \times n$ matrix, and let the $(i, j)^{\text{th}}$ entry of $G(p)$ be denoted as g_{ij} (Note that the matrix representation of G is assumed to be in some fixed frame bundle basis). Also, denote the $(i, j)^{\text{th}}$ entry of $G(p)^{-1}$ as g^{ij}. Then we define:

$$\Gamma_{ij}^{k} := \frac{1}{2} \sum_{\ell=1}^{n} g^{\ell k} \cdot \left(\frac{\partial g_{\ell i}}{\partial y_j}(y) + \frac{\partial g_{\ell j}}{\partial y_i}(y) - \frac{\partial g_{ij}}{\partial y_\ell}(y) \right)$$

Length and Energy

Let $\gamma : [0, 1] \mapsto (M, g)$ be a smooth map. We define the length of γ as:

$$L(\gamma) := \int_0^1 \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{g(\gamma(t))}} dt \tag{1}$$

We also define the *energy* of γ as:

$$\mathcal{E}(\gamma, \dot{\gamma}, t) := \frac{1}{2} \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{g(\gamma(t))} \tag{2}$$

Geodesics

Definition (Geodesics)

Define the energy functional to be:

$$S(\gamma) := \int_0^1 \mathcal{E}(\gamma, \dot{\gamma}, t)dt = \frac{1}{2}\int_0^1 \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{g(\gamma(t))} dt$$

Effectively, $S(\gamma)$ denotes the work done to move the particle along the curve.
A curve $\gamma_* : [0, 1] \mapsto M$ is called a *geodesic* if it is a critical point of the energy functional.

A geodesic may not always exist: For example, consider the manifold $\mathbb{R}^2 - \{(0, 0)\}$ equipped with the usual metric tensor. There is no geodesic between $(1, 1)$ and $(-1, -1)$ on this manifold. Thus, from this point onwards, we shall only deal with Riemannian manifolds for which there is a geodesic between any two given points.

Geodesic Equation

Using the Euler-Lagrange equations to characterize geodesics yields that γ_* must satisfy the following differential equation:

$$\frac{d}{dt}\frac{\partial \mathcal{E}}{\partial \dot{\gamma}} = \frac{\partial \mathcal{E}}{\partial \gamma} \tag{3}$$

Simplifying the above, we obtain that a geodesic γ_* must satisfy the following series of second-order differential equations: Indeed, fix a point $p \in M$, and consider a chart (U, φ) such that $p \in U$. Define $x_* : [0, 1] \mapsto \mathbb{R}^n$ as $x_* := \varphi \circ \gamma_*$. Then, for every $k \in [n]$, we have:

$$(\ddot{x}_*)_k = -\sum_{i,j=1}^{n} \Gamma_{ij}^{k}(x_*)(\dot{x}_*)_i(\dot{x}_*)_j \tag{4}$$

Geodesics for various Manifolds

Consider the manifold $\mathbb{R}^n_{>0}$, whose smoothness is inherited from $\mathbb{R}^n$. Consider the *log-barrier* function $f : \mathbb{R}^n_{>0} \mapsto \mathbb{R}$, given by $f(x_1, \ldots, x_n) := -\sum_{i=1}^n \ln(x_i)$. Then the Hessian of f is a diagonal matrix whose diagonal entries are $x_1^{-2}, \ldots, x_n^{-2}$. Thus, for any point $p = (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}$, we define $G(p) := \operatorname{diag}(x_1^{-2}, \ldots, x_n^{-2})$. It is not too difficult to see that the Christoffel symbols are given by

$$\Gamma_{ij}^k = \begin{cases} -1/x_k & \text{if } i = j = k \\ 0 & \text{otherwise} \end{cases}$$

Eq. (4) then simplifies to $(\ddot{x})_k = (\dot{x})_k^2 / x_k$ for every $k \in [n]$, solving which yields $x(t) = c_1 \cdot c_2^t$ for some constants $c_1, c_2 \in \mathbb{R}^n_{>0}$, where c_2^t is defined componentwise. Now, assuming $x(0) = p, x(1) = q$, we obtain the geodesic on the positive orthant between p and q is parametrized as:

$$x(t) = (p_1(q_1/p_1)^t, \ldots, p_n(q_n/p_n)^t)$$

Geodesics in $\mathbb{S}^k_{++}$

Let $\mathbb{S}^k_{++}$ be the space of all positive-**definite** symmetric $k \times k$ matrices, considered as a submanifold of $\mathbb{R}^{k(k+1)/2}$. A metric tensor on $\mathbb{S}^k_{++}$ at some point $S \in \mathbb{S}^k_{++}$ is given by $g_S(U, V) := \mathrm{tr}(S^{-1}US^{-1}V)$.
We shall solve the Euler-Lagrange equations directly for this manifold. Before we begin with that, define:

$$E_{ij} = \begin{cases} e_i e_j^\mathsf{T} + e_j e_i^\mathsf{T} & \text{if } i \neq j \\ e_i e_i^\mathsf{T} & \text{otherwise} \end{cases}$$

Note that $\{E_{ij}\}_{i,j\in[k]}$ form a basis for $\mathcal{S}_k$, where $\mathcal{S}_k$ is the set of *all* symmetric matrices. Thus we can write $\gamma : [0, 1] \mapsto \mathbb{S}^k_{++} \hookrightarrow \mathcal{S}_k$ as

$$\gamma(t) = \sum_{i,j\in[k]} \gamma_{ij}(t) E_{ij}$$

Consequently,

$$\frac{\partial \gamma}{\partial \gamma_{ij}} = E_{ij}$$

Now, from Eq. (2) (we ignore the factor of $1/2$),

$$\mathcal{E}(\gamma, \dot{\gamma}, t) = \operatorname{tr}(\gamma(t)^{-1}\dot{\gamma}(t)\gamma(t)^{-1}\dot{\gamma}(t))$$

Thus,

$$\frac{\partial \mathcal{E}}{\partial \gamma_{ij}} = \operatorname{tr}\left(\frac{\partial \gamma^{-1}}{\partial \gamma_{ij}}\dot{\gamma}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\frac{\partial \dot{\gamma}}{\partial \gamma_{ij}}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\dot{\gamma}\frac{\partial \gamma^{-1}}{\partial \gamma_{ij}}\dot{\gamma} + \gamma^{-1}\dot{\gamma}\gamma^{-1}\frac{\partial \dot{\gamma}}{\partial \gamma_{ij}}\right)$$

Also, we have

$$\frac{\partial \gamma^{-1}}{\partial \gamma_{ij}} = -\gamma^{-1}E_{ij}\gamma^{-1}$$

$$\frac{\partial \dot{\gamma}}{\partial \gamma_{ij}} = 0$$

Thus,

$$\frac{\partial \mathcal{E}}{\partial \gamma_{ij}} = -2\operatorname{tr}\left(\gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\right) = -2\operatorname{tr}\left(E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1}\right)$$

Similarly,

$$\frac{\partial \mathcal{E}}{\partial \dot{\gamma}_{ij}} = \operatorname{tr}\left(\frac{\partial \gamma^{-1}}{\partial \dot{\gamma}_{ij}}\dot{\gamma}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\frac{\partial \dot{\gamma}}{\partial \dot{\gamma}_{ij}}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\dot{\gamma}\frac{\partial \gamma^{-1}}{\partial \dot{\gamma}_{ij}}\dot{\gamma} + \gamma^{-1}\dot{\gamma}\gamma^{-1}\frac{\partial \dot{\gamma}}{\partial \dot{\gamma}_{ij}}\right)$$

$$= \operatorname{tr}\left(\gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\dot{\gamma}\gamma^{-1}E_{ij}\right)$$

But $\operatorname{tr}\left((\gamma^{-1}\dot{\gamma}) \cdot (\gamma^{-1}E_{ij})\right) = \operatorname{tr}\left((\gamma^{-1}E_{ij}) \cdot (\gamma^{-1}\dot{\gamma})\right)$, and thus

$$\frac{\partial \mathcal{E}}{\partial \dot{\gamma}_{ij}} = 2\operatorname{tr}\left(\gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma}\right)$$

Now,

$$\frac{d}{dt}\frac{\partial \mathcal{E}}{\partial \dot{\gamma}_{ij}}$$

$$= 2\operatorname{tr}\left(\frac{d\gamma^{-1}}{dt}E_{ij}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\frac{dE_{ij}}{dt}\gamma^{-1}\dot{\gamma} + \gamma^{-1}E_{ij}\frac{d\gamma^{-1}}{dt}\dot{\gamma} + \gamma^{-1}E_{ij}\gamma^{-1}\frac{d\dot{\gamma}}{dt}\right)$$

Once again,

$$\frac{d\gamma^{-1}}{dt} = -\gamma^{-1}\dot{\gamma}\gamma^{-1}$$

Thus

$$\frac{d}{dt}\frac{\partial \mathcal{E}}{\partial \dot{\gamma}_{ij}} = 2\,\mathrm{tr}\left(-\gamma^{-1}\dot{\gamma}\gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma} - \gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma} + \gamma^{-1}E_{ij}\gamma^{-1}\ddot{\gamma}\right)$$

$$= 2\,\mathrm{tr}\left(-E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} + E_{ij}\gamma^{-1}\ddot{\gamma}\gamma^{-1}\right)$$

$$= 2\,\mathrm{tr}\left(E_{ij}(\gamma^{-1}\ddot{\gamma}\gamma^{-1} - 2\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1})\right)$$

Thus, applying the Euler-Lagrange equations (Eq. (3)) yield:

$$-2\,\mathrm{tr}\left(E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1}\right) = 2\,\mathrm{tr}\left(E_{ij}(\gamma^{-1}\ddot{\gamma}\gamma^{-1} - 2\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1})\right)$$

$$\implies \mathrm{tr}\left(E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1}\right) = \mathrm{tr}\left(E_{ij}(-\gamma^{-1}\ddot{\gamma}\gamma^{-1} + 2\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1})\right)$$

$$\implies \operatorname{tr}\left(E_{ij}(\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1})\right) = 0$$

But

$$\operatorname{tr}\left(E_{ij}(\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1})\right) = \langle E_{ij}, \gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1}\rangle_{\text{Frob}}$$

Since $\{E_{ij}\}$'s span $\mathcal{S}_k$,

$$\langle S, \gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1}\rangle_{\text{Frob}} = 0$$

for all $S \in \mathcal{S}_k$, and consequently,

$$\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1} = 0 \implies \ddot{\gamma}\gamma^{-1} - \dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} = 0$$

$$\implies \frac{d(\dot{\gamma}\gamma^{-1})}{dt} = 0$$

Consequently,

$$\dot{\gamma}\gamma^{-1} = C \implies \dot{\gamma} = C\gamma$$

This is a matrix-valued differential equation with the solution $\gamma(t) = \exp(tC)\gamma(0)$. Now, suppose we want to find geodesic(s) joining matrices $P, Q \in \mathbb{S}^k_{++}$. Then $\gamma(0) = P$. We still need to find C such that $\gamma(1) = Q$. To that extent, we perform a 'diagonalization' trick, where we write $C = P^{1/2}SP^{-1/2}$. Note that $C^\ell = P^{1/2}S^\ell P^{-1/2}$, for any $\ell \in \mathbb{N}_0$. Then,

$$\gamma(t) = \exp(tC)P = \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} C^\ell P = \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} P^{1/2}S^\ell P^{-1/2}P = P^{1/2}\exp(tS)P^{1/2}$$

Thus

$$\gamma(1) = Q \implies \exp(S) = P^{-1/2}QP^{-1/2}$$

Now, since $S \in \mathbb{S}^k_{++}$, $S = UD'U^\mathsf{T}$ for some diagonal matrix D'. Then

$$\exp(S) = \sum_{\ell=0}^{\infty} \frac{S^\ell}{\ell!} = \sum_{\ell=0}^{\infty} U\frac{D'^\ell}{\ell!}U^\mathsf{T} = U\exp(D')U^\mathsf{T}$$

Thus,

$$\gamma(t) = P^{1/2}\exp(tS)P^{1/2} = P^{1/2}(\exp(S))^t P^{1/2} = P^{1/2}(P^{-1/2}QP^{-1/2})^t P^{1/2}$$

Consequently, on $\mathbb{S}^k_{++}$, a geodesic between any two points P, Q is parametrized as $P^{1/2}(P^{-1/2}QP^{-1/2})^t P^{1/2}$.

We now change tracks and define geodesic convexity.

Definition (Geodesic Convexity)

A subset $S \subset M$ of a Riemannian manifold M is geodesically convex if for all $x, y \in S$, there exists a geodesic $\gamma : [0,1] \mapsto M, \gamma(0) = x, \gamma(1) = y$, such that $\gamma([0,1]) \subset S$.

Definition (Geodesic Convexity of functions)

Let S be a geodesically convex subset of some ambient manifold M. A function $f : S \mapsto \mathbb{R}$ (not necessarily continuous/smooth) is called geodesically (strictly) convex if $f \circ \gamma : [0,1] \mapsto \mathbb{R}$ is (strictly) convex for all geodesics $\gamma : [0,1] \mapsto M$ such that $\gamma(0) \neq \gamma(1)$ and $\gamma([0,1]) \subset S$. In other words, for any geodesic γ, we have

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)), t \in [0,1]$$

Definition (Geodesic Strong Convexity)

Let S be a geodesically convex subset of some manifold M. A function $f : S \mapsto \mathbb{R}$ is called geodesically μ-strongly convex if for all geodesics $\gamma : [0, 1] \mapsto M$ with $\gamma([0, 1]) \subset S$, we have:

$$f(\gamma(t)) \leq (1 - t)f(\gamma(0)) + tf(\gamma(1)) - \frac{t(1-t)\mu}{2} L(\gamma)^2$$

where $L(\gamma)$ is the length of γ (recall Eq. (1)). Equivalently, f is geodesically μ-strongly convex if $f \circ \gamma : [0, 1] \mapsto \mathbb{R}$ is $\mu L(\gamma)$-strongly convex.

Definition (Concavity and Linearity)

Let S be a geodesically convex subset of some ambient manifold M. A function $f : S \mapsto \mathbb{R}$ is called geodesically concave if $-f$ is geodesically convex, and f is called geodesically linear if f is both geodesically convex and concave.

Since geodesic convexity has been defined by pulling back along geodesics and then imposing usual convexity, many of the usual properties of convexity transfer through.

Lemma (Local Minimizers are Global Minimizers)

If $f : S \mapsto \mathbb{R}$ is geodesically convex, then any local minimizer is a global minimizer. In particular, if $f : S \mapsto \mathbb{R}$ is geodesically strictly convex, then it has at most one local minimizer, which, by the above lemma, must also be a global minimizer.

Differentiable Convex Functions

We eventually hope to build a theory of optimization of geodesically convex functions. For such purposes, having some differentiability helps.

Theorem (Differentiability and Convexity)

Let S be a geodesically convex set, and let $f : M \mapsto \mathbb{R}$ be differentiable in a neighborhood of S. Then $f|_S$ is geodesically convex if and only if for every geodesic $\gamma : [0, 1] \mapsto M, \gamma([0, 1]) \subset S$, we have:

$$f(\gamma(t)) \geq f(\gamma(0)) + t\langle \nabla f, \gamma'(0)\rangle_{g(\gamma(0))}, \forall t \in [0, 1]$$

$f|_S$ is geodesically μ-strongly convex if and only if

$$f(\gamma(t)) \geq f(\gamma(0)) + t\langle \nabla f, \gamma'(0)\rangle_{g(\gamma(0))} + \frac{\mu t^2}{2} L(\gamma)^2, \forall t \in [0, 1] \quad (5)$$

Corollary

If $f : S \mapsto \mathbb{R}$ is differentiable and geodesically convex, where S is open and geodesically convex, then x is a global minimizer of f if and only if $(\nabla f)(x) = 0$.

We now also give characterizations based on second-order derivatives, without proof (the proof is essentially identical to that in standard convex analysis, except for the fact that we have to pull f back along the geodesics):

Theorem

Let $f : S \mapsto \mathbb{R}$ be twice-differentiable, and assume S is open and geodesically convex. Then f is:

1. *Geodesically convex if and only if* $\operatorname{Hess} f(x) \succeq 0$.
2. *Geodesically μ-strongly convex if and only if* $\operatorname{Hess} f(x) \succeq \mu \cdot \mathrm{Id}$.
3. *Geodesically strictly convex if* $\operatorname{Hess} f(x) \succ 0$.

Note that checking geodesic convexity of f eventually boils down to checking convexity of $f \circ \gamma$ in $\mathbb{R}^n$. Thus, using the formulae for geodesics in $\mathbb{R}^n_{>0}$ (with log-barrier induced metric tensor) and $\mathbb{S}^k_{++}$, we have the following results.

Lemma (Geodesic Linearity of the log-barrier function)

The map $\mathbb{R}^n_{>0} \ni x \mapsto \langle 1, \ln(x) \rangle \in \mathbb{R}$ *is geodesically linear on the manifold* $\mathbb{R}^n_{>0}$.

Lemma (Geodesic Log-Convexity of Polynomials with positive coefficients)

Let $p(x_1, \ldots, x_n)$ *be a multivariate polynomial with positive coefficients. Then* p *is geodesically* ***log****-convex on* $\mathbb{R}^n_{>0}$*, i.e.* $\ln p(\gamma(\cdot))$ *is convex in* $\mathbb{R}^n$.

We shall now give analogs of all the above examples for the manifold $\mathbb{S}^n_{++}$. There is indeed an analogy between $\mathbb{R}^n_{>0}$ and $\mathbb{S}^n_{++}$ in the sense that they are both Hessian manifolds, that too of "log-barrier functions" ($\sum \ln(x_i)$ for $\mathbb{R}^n_{>0}$, $\ln\det(X)$ for $\mathbb{S}^n_{++}$).

Lemma (Geodesic Linearity of the log-det map)

The map $\mathbb{S}^n_{++} \ni X \mapsto \ln\det(X) \in \mathbb{R}$ is geodesically linear on the manifold $\mathbb{S}^n_{++}$ as defined in the first chapter.

Lemma (Geodesic Convexity of Strictly Positive Linear Operators)

Let $T : \mathcal{S}_n \mapsto \mathcal{S}_m$ be a linear map such that $T(\mathbb{S}^n_{++}) \subseteq \mathbb{S}^m_{++}$. Such maps are also called strictly positive linear operators. Then $\mathbb{S}^n_{++} \ni X \mapsto T(X) \in \mathbb{S}^m_{++}$ is a geodesically convex map w.r.t the order $\preceq$ on $\mathbb{S}^m_{++}$, i.e. for any geodesic $\gamma : [0,1] \mapsto \mathbb{S}^n_{++}$, and any $t \in [0,1]$, we have

$$T(\gamma(t)) \preceq (1-t)T(\gamma(0)) + tT(\gamma(1))$$

Lemma (Geodesic Convexity of log-det of Strictly Positive Operators)

Let $T : \mathcal{S}_n \mapsto \mathcal{S}_m$ *be a linear map such that* $T(\mathbb{S}^n_{++}) \subseteq \mathbb{S}^m_{++}$*. Then* $\mathbb{S}^n_{++} \ni X \mapsto \ln\det(T(X)) \in \mathbb{R}$ *is a geodesically convex map.*

Before we get to Gradient Descent on Riemannian manifolds, we rephrase Eq. (5) as:

$$f(\exp_x(tv)) \geq f(x) + t\langle \nabla f(x), v\rangle_x + \frac{\mu t^2}{2} \cdot \|v\|_x^2 \tag{6}$$

where we use the fact that $L(\gamma_{x,v}|_{[0,1]}) = \|v\|_x =: \sqrt{\langle v, v\rangle_x}$.
The way we use this inequality for optimization is by topping it off with an upper bound: Indeed, if ∇f is L-Lipschitz continuous, then

$$f(\exp_x(tv)) \leq f(x) + t\langle \nabla f(x), v\rangle_x + \frac{Lt^2}{2} \cdot \|v\|_x^2 \tag{7}$$

Lemma (Polyak-Łojasiewicz Inequality)

Let S be a non-empty, closed, and geodesically convex subset of M, where M is a complete Riemannian manifold. Assume $f : M \mapsto \mathbb{R}$ is differentiable on a neighborhood of S. If $f|_S$ is geodesically μ-strongly convex with $\mu > 0$, then

$$f(x) - f(x_*) \leq \frac{1}{2\mu} \|\nabla f(x)\|_x^2$$

for all $x \in S$, where x_* is the unique global minimizer of $f|_S$.

Theorem (Riemannian Gradient Descent)

Let $f : M \mapsto \mathbb{R}$ be a differentiable geodesically convex function on a complete connected manifold M. Let $x_0 \in M$, and consider the sublevel set $S_0 := \{x \in M : f(x) \leq f(x_0)\}$. Assume f has a L-Lipschitz continuous gradient on a neighborhood of S_0, and suppose $f|_{S_0}$ is geodesically μ-strongly convex. Consider ***gradient descent with exponential retraction*** *and step-size $1/L$ initialized at x_0, i.e.*

$$x_{k+1} = \exp_{x_k}\left(-\frac{1}{L}\nabla f(x_k)\right), k \in \mathbb{N}_0$$

There exists a unique global minimizer of $f|_{S_0}$ which is $x_* \in S_0$, and convergence to x_* is linear, i.e. if we set $\kappa = L/\mu$, then we have that $x_k \in S_0$ for all $k \in \mathbb{N}_0$, and

$$\frac{f(x_k) - f(x_*)}{f(x_0) - f(x_*)} \leq \left(1 - \frac{1}{\kappa}\right)^k, \frac{d(x_k, x_*)}{d(x_0, x_*)} \leq \sqrt{\kappa} \cdot \left(1 - \frac{1}{\kappa}\right)^{k/2}$$

Applications

Before we describe the applications of geodesic convexity, we take a brief detour into functional analysis and state the very important **Brascamp-Lieb inequality**:

Theorem (Brascamp-Lieb Inequality)

Given linear maps $B = (B_j)_{j\in[m]}$, $B_j : \mathbb{R}^n \mapsto \mathbb{R}^{n_j}$, *and non-negative real numbers* $(p_j)_{j\in[m]}$, *there exists a number* $C \in [0, \infty]$ *such that for any tuple of measurable functions* $(f_j)_{j\in[m]}$, $f : \mathbb{R}^{n_j} \mapsto \mathbb{R}_{\geq 0}$, *the following inequality holds:*

$$\int_{x\in\mathbb{R}^n} \prod_{j=1}^{m} f_j(B_j x)^{p_j} dx \leq C \prod_{j=1}^{m} \left(\int_{x\in\mathbb{R}^{n_j}} f_j(x) dx \right)^{p_j}$$

The smallest C for which the above inequality holds is called the *Brascamp-Lieb constant* for the system (B, p), and is denoted as $\mathrm{BL}(B, p)$. A system (B, p) is called *feasible* if $\mathrm{BL}(B, p) < \infty$.

Bennett, Carbery, Christ, and Tao showed that $\mathrm{BL}(B, p) < \infty$ if and only if the following criteria are satisfied:

1. $n = \sum_{j \in [m]} p_j n_j$.
2. $\dim(V) \leq \sum_{j \in [m]} p_j \dim(B_j V)$ for any subspace V of $\mathbb{R}^n$.

Henceforth, we will only be working with feasible Brascamp-Lieb systems.

Now, Lieb showed that equality occurs in the Brascamp-Lieb inequality when $f_j(x) = \exp(-x^\mathsf{T} A_j x)$ for some positive definite matrix A_j, for all $j \in [m]$. Plugging the above into the Brascamp-Lieb inequality yields:

$$\mathrm{BL}(B, p) \geq \left(\frac{\prod_{j \in [m]} \det(A_j)^{p_j}}{\det\left(\sum_{j \in [m]} p_j B_j^\mathsf{T} A_j B_j \right)} \right)^{1/2}$$

And thus

$$\mathrm{BL}(B, p) = \sup_{(X_1, \ldots, X_m)} \left(\frac{\prod_{j \in [m]} \det(X_j)^{p_j}}{\det\left(\sum_{j \in [m]} p_j B_j^\mathsf{T} X_j B_j \right)} \right)^{1/2}$$

where $X_j \in \mathbb{S}_{++}^{n_j}$ for all $j \in [m]$.

To simplify the expression a bit, we usually deal with the negative logarithm of it. Under that, we obtain:

$$-\ln \mathrm{BL}(B, p) = -\frac{1}{2} \sup_{(X_1,\ldots,X_m)} \left(\sum_{j\in[m]} p_j \ln\det(X_j) - \ln\det\left(\sum_{j\in[m]} p_j B_j^\top X_j B_j \right) \right)$$

Now, it can be shown that the function of $(X_1, \ldots, X_m)$ inside the supremum is not concave in the usual Euclidean sense: Indeed, suppose it was. Fix $X_2, \ldots, X_m$. Then we are effectively dealing with

$$p_1 \ln\det(X_1) - \ln\det\left(p_1 B_1^\top X_1 B_1 + C\right)$$

where C is some positive definite matrix. Note that $X_1 \mapsto \ln\det(X_1)$, and $X_1 \mapsto \ln\det\left(p_1 B_1^\top X_1 B_1 + C\right)$ are both concave functions and thus it is difficult to comment on the concavity of their difference. Indeed, there exist values of $X_2, \ldots, X_m$ for which the above function is not concave.

Thus, the usual tools of convex optimization fail for this problem. Here comes the true power of geodesics: We will prove that the above formulation is geodesically concave.
Before coming to analyses of convexity, we state another equivalent characterization of the Brascamp-Lieb constant:

$$-2 \ln \mathrm{BL}(B, p) = \inf_{X \in \mathbb{S}^n_{++}} F_{B,p}(X)$$

where

$$F_{B,p}(X) := \sum_{j \in [m]} p_j \ln \det(B_j X B_j^{\mathsf{T}}) - \ln \det(X)$$

Theorem

$F_{B,p}$ is a geodesically convex function on the manifold $\mathbb{S}^n_{++}$ with the dot product induced by the Riemannian tensor being given by $g_X(U, V) := \mathrm{tr}(X^{-1} U X^{-1} V)$.

Acknowledgements

I thank Prof. Debasish Chatterjee for suggesting the topic of this presentation to me, and giving his time to discuss things with me.

The End

Questions? Comments?