GEODESIC
CONVEXITY

Nisheeth Vishnoi

https://nisheethvishnoi.wordpress.com/convex-optimization/

Institute for Advanced Study, June 7, 2018
Convexity and Optimization

For a convex function local minimum = global min.

\[ tf(x) + (1 - t)f(y) \geq f(tx + (1 - t)y) \]

\[ f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \nabla^2 f(x) \succeq 0 \]

Goal: \( f(\hat{x}) \leq f(x^*) + \varepsilon \)

Gradient Descent
\[
\frac{dx}{dt} = -\nabla f(x)
\]
\[ x^{k+1} = x^k - \nabla f(x^k) \]

Newton-type methods
\[
\frac{dx}{dt} = -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x)
\]
\[ x^{k+1} = x^k - \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k) \]

Cutting plane methods
more later ...

Running Times: \( T(n, \varepsilon, \|\nabla f\|, \kappa(\nabla^2 f), \|x^0 - x^*\|, t_{\text{grad}}, t_{\text{Hessian}}, \ldots) \)

Ushered in an TCS/ML revolution ...
Optimization Problems: Commutative and Non-Commutative

Given evaluation oracle to $p(x) \in \mathbb{R}_+[x_1, ..., x_m]$ and $\theta \in \mathbb{R}^m$

\[ P1: \inf_{x \in \mathbb{R}^m_+} \log p(x) - \sum \theta_i \log x_i \]

Applications to discrete counting problems [Gurvits ’04, SinghV. ‘14, StraszakV. ’17a]

\[ \log p(x) \] is not convex (sometimes concave) — not a convex optimization problem!

Given $m \ell \times n$ real-valued matrices $B_1, B_2, ..., B_m$ and a $\theta \in \mathbb{R}^m_+$

\[ P2: \inf_{X > 0} \sum \theta_j \log \det (B_j X B_j^\top) - \log \det X \]

Applications to Brascamp-Lieb const. [SraV.Yildiz ‘18]; studied by [BCCT ’05, GGOW+ ‘16+]

Not a convex optimization problem either (rank 1 case as above)

Both problems are geodesically convex!
## Convexity vs Geodesic Convexity

<table>
<thead>
<tr>
<th>Euclidean space</th>
<th>Smooth manifolds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculus (differentiation / integration)</td>
<td>Affine connections</td>
</tr>
<tr>
<td>Straight Lines</td>
<td>Geodesics</td>
</tr>
<tr>
<td>Convex Sets</td>
<td>Geodesically convex sets</td>
</tr>
<tr>
<td>Convex functions</td>
<td>Geodesically convex functions</td>
</tr>
<tr>
<td>Local = Global</td>
<td>Local = global</td>
</tr>
</tbody>
</table>

### Plan for the talk:

- a) Manifolds, Geodesics, Geodesic convexity
- b) Geodesic convexity of the applications
- c) An algorithm for P1

---

**Plan for the talk:**

- a) Manifolds, Geodesics, Geodesic convexity
- b) Geodesic convexity of the applications
- c) An algorithm for P1
MANIFOLDS, GEODESICS, GEODESIC CONVEXITY
**Manifolds, Calculus and Metrics**

Smooth manifolds
\( \mathcal{X}(M) : \) vector fields over \( M \)

Curves

**Euclidean Space:** \( D_X(f_1, ... , f_n) \) is just the directional derivative

<table>
<thead>
<tr>
<th>Affine Connection: ( \nabla: \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{X} )</th>
<th>Riemannian Metric Tensor: ( g_p(u,v) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \forall X,Y,Z \in \mathcal{X}(M), \quad \forall f \text{ on } M )</td>
<td>( \forall u,u',v \in T_pM, \forall c \in \mathbb{R} )</td>
</tr>
<tr>
<td>Linear in first term: ( \nabla_{X+fY}Z = \nabla_XZ + f\nabla_YZ )</td>
<td>Symmetric: ( g_p(u,v) = g_p(v,u) )</td>
</tr>
<tr>
<td>Linear in second term: ( \nabla_XY + Z = \nabla_XY + \nabla_XZ )</td>
<td>Bilinear: ( g_p(u + cu', v) = g_p(u,v) + cg_p(u',v) )</td>
</tr>
<tr>
<td>Leibniz’s rule: ( \nabla_XfY = f\nabla_XY + YD_Xf )</td>
<td>Positive definite: ( g_p(u,u) &gt; 0, u \neq 0 )</td>
</tr>
</tbody>
</table>

Compatibility: \( D_X(g(Y,Z)) = g(\nabla_XY, Z) + g(Y, \nabla_XZ) \)

Torsion free: \( \nabla_XY - \nabla_YX = [X,Y] \)

**Levi-Civita connection:** Unique, torsion-free, affine-connection compatible with metric
Examples

**Manifold:** Positive Orthant $\mathbb{R}^m_+$

**Tangent Space:** $\mathbb{R}^m$

**Riemannian Metric:** For $p \in \mathbb{R}^m_+$, and $u, v \in \mathbb{R}^m$

$$g(u, v) := \langle P^{-1}u, P^{-1}v \rangle = \sum \frac{u_i v_i}{p_i^2}$$

**Hessian:** Let $f(p) = -\sum \log p_i$

Then $g = \text{Hessian of } f$

**Levi-Civita Connection:** At a point $p$

$$\nabla_{e_i} e_i = p_i^{-1}e_i = 0 \text{ o.w.}$$

---

**Manifold:** Positive Definite Matrices $\mathbb{S}^n_{++}$

**Tangent Space:** $\mathbb{S}^n$

**Riemannian Metric:** For $P \in \mathbb{S}^n_{++}$, and $U, V \in \mathbb{S}^n$

$$g(U, V) := \text{Tr } P^{-1}UP^{-1}V$$

**Hessian:** Let $f(P) = -\log \det P$

Then $g = \text{Hessian of } f$

Calculations of Levi-Civita get more complicated …
Geodesics: Two Views

Curves that take tangent vectors "parallel" on the curve

\[ \dot{\gamma} = q - p \]

\[ \ddot{\gamma} = 0 \]

\[ \nabla_{\dot{\gamma}} \dot{\gamma} = 0 \]

Shortest curves between points

\[ S[\nu] = \int g_{\nu}(\dot{\nu}, \dot{\nu}) dt \]

\[ \gamma = \operatorname{arginf}_\nu S[\nu] \]

\[ \ddot{\gamma} = 0 \]

Euler-Lagrange Dynamics
Examples

**Manifold:** Positive Orthant $\mathbb{R}^m_+$

**Levi-Civita Connection:** At a point $p$

\[ \nabla_{e_i} e_i = p_i^{-1} e_i \]

**Geodesic Equation:** $\nabla \dot{\gamma} \dot{\gamma} = 0$

**Simplify (exercise):**

\[ \forall i \quad \ddot{\gamma}_i = \dot{\gamma}_i^2 \gamma_i^{-1} \]

**Solve ODE:**

\[ \frac{d}{dt} \log \dot{\gamma}_i = \frac{d}{dt} \log \gamma_i \]

\[ \dot{\gamma}_i = \alpha_i \gamma_i \quad \text{for some } \alpha_i \]

\[ \gamma_i = \beta_i e^{\alpha_i t} \quad \text{for some } \beta_i > 0, \alpha_i \]

---

**Manifold:** Positive Definite Matrices $\mathbb{S}^n_{++}$

**Geodesic Equation:**

\[ \frac{\partial g_\gamma (\dot{\gamma}, \dot{\gamma})}{\partial \gamma} \frac{d}{dt} \frac{\partial g_\gamma (\dot{\gamma}, \dot{\gamma})}{\partial \dot{\gamma}} = \frac{d}{dt} \frac{\partial g_\gamma (\dot{\gamma}, \dot{\gamma})}{\partial \dot{\gamma}} \]

**Simplify:**

\[ \dot{\gamma} \gamma^{-1} = C \quad \text{for some constant matrix } C \]

**Solve:**

\[ \gamma(t) = \exp(tC) D \]

**Geodesic between $P, Q$:**

\[ P^\frac{1}{2} \left( P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right)^t P^\frac{1}{2} \]
Geodesics Convexity

**0th Order Characterization**
- \( f : M \to \mathbb{R} \) is geodesically convex if for any geodesic \( \gamma : [0,1] \to M \) and \( \forall t \in [0,1] \)
  - \( f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)) \)

**1st Order Characterization**
- \( f : M \to \mathbb{R} \) is geodesically convex iff for any \( p, q \in M \) with geodesic joining them \( \gamma_{pq} \),
  - \( f(p) + \dot{\gamma}_{pq}(f)(p) \leq f(q) \)

**2nd Order Characterization**
- \( f : M \to \mathbb{R} \) is geodesically convex if for any geodesic \( \gamma : [0,1] \to M \) and \( \forall t \in [0,1] \)
  - \( \frac{d^2 f(\gamma(t))}{dt^2} \geq 0 \)

If \( f \) is geodescially convex, then every local minimum is also a global minimum.
Geodesic Convexity vs Non-convexity

\[ \ln(2)(x - 2) + \ln(2)^2 \]

\[ \ln(2)^2 + 2 \ln(2) \ln \left( \frac{x}{2} \right) \]

\[ e^x \sin x \]
Theorem: Given $p(x) = \sum_{\tau \in \mathcal{F}} c_\tau x^\tau \in \mathbb{R}_+[x_1, \ldots, x_m]$ where $x^\tau = \prod_{j \in [m]} x_j^{\tau_j}$ and $\mathcal{F} \subset \mathbb{Z}_{\geq 0}^m$, $\log p(x)$ is geodesically convex.

Geodesic: $\gamma(t) := (\beta_1 e^{t\alpha_1}, \ldots, \beta_m e^{t\alpha_m})$ for real vectors $\beta \in \mathbb{R}_+^m$ and $\alpha \in \mathbb{R}^m$.

Second Order Convexity: $\log p(x)$ is geodesically convex if for any geodesic $\gamma(t)$

$$\forall t \in [0,1], \quad \frac{d^2 \log p(\gamma(t))}{dt^2} \geq 0$$

First derivative:

$$\frac{d \log p(\gamma(t))}{dt} = \frac{\dot{p}}{p} = \frac{\sum_{\tau \in \mathcal{F}} c_\tau \langle \alpha, \tau \rangle \gamma(t)^\tau}{\sum_{\tau \in \mathcal{F}} c_\tau \gamma(t)^\tau}$$

Second derivative:

$$\frac{d^2 \log p(\gamma(t))}{dt^2} = \frac{\ddot{p}}{p} - \left(\frac{\dot{p}}{p}\right)^2 = \frac{\sum_{\tau, \tau' \in \mathcal{F}} (\langle \alpha, \tau \rangle - \langle \alpha, \tau' \rangle)^2 c_\tau c_{\tau'} \gamma(t)^\tau \gamma(t)^{\tau'}}{\left(\sum_{\tau \in \mathcal{F}} c_\tau \gamma(t)^\tau\right)^2} \geq 0$$
Geodesic Convexity of Brascamp-Lieb

Given \( m \times n \) real-valued matrices \( B_1, B_2, ..., B_m \) and a \( \theta \in \mathbb{R}^m_+ \)

\[
\inf_{X > 0} \sum \theta_j \log \det (B_j X B_j^T) - \log \det X
\]

\textbf{Theorem(s) [SraV.Yildiz ‘18]:} Geodesically convex and computes BL-constant!

\textbf{Geodesic:} Given PD matrices \( P \) and \( Q \), the geodesic between them

\[
\gamma(t) := P^{1/2} \left( P^{-1/2} Q P^{-1/2} \right)^t P^{1/2}
\]

\textbf{Simple Fact:} \( \log \det X \) is geodesically linear

\[
\forall t \in [0,1], \quad \log \det(\gamma(t)) = (1 - t) \log \det P + t \log \det Q
\]

\textbf{Theorem [AndoKubo ‘79]:} If \( T(X) \) is a strictly positive linear operator, then \( \log \det T(X) \) is geodesically convex

\textbf{By taking positive combinations, enough to show:} \( T_j(X) = B_j X B_j^T \) is a strictly positive linear map for \( j \in [m] \) if \( \ell \sum \theta_j = n \) and \( \dim(\mathbb{R}^n) = \sum_{j \in [m]} \theta_j \dim(B_j \mathbb{R}^n) \)

\textbf{Proof:} Assume \( T_i(X) \) is not strictly positive linear. Then for some \( X \in \mathbb{S}^n_{++} \), there exists \( v \in \mathbb{R}^\ell \) such that \( v^T T_i(X) v \leq 0 \). Thus, \( (B_i^T v)^T X (B_i^T v) \leq 0 \). Thus, \( (B_i^T v) = 0 \) and \( \dim(B_i \mathbb{R}^n) < \ell \).

Consequently

\[
n = \dim(\mathbb{R}^n) = \sum_{j \in [m]} \theta_j \dim(B_j \mathbb{R}^n) < \sum_{j \in [m]} \theta_j \ell = n - \text{contradiction!}
\]
ALGORITHM FOR P1
(RANK ONE BL)
Ellipsoid Method

\[
\text{OPT} = \inf_{y \in \mathbb{R}^m} \log p(e^y) - \sum \theta_i y_i = \inf_{y \in \mathbb{R}^m} f(y)
\]

Reduce to Feasibility: Given $A$, check if OPT is $\leq A + \varepsilon$ or $> A$

Assume $\|y^*\| \leq R, f \in [-M, M]$

Ellipsoid Algorithm:

- **Start** with an ellipsoid $E_0$ that contains $y^*$
- At $k$th step, let $E_k$ be the ellipsoid centered at $y^k$
  - IF $f(y^k) \leq A$, DONE
  - ELSE
    - use evaluation oracle for $p$ to get $\nabla f(y^k)$
    - $E_{k+1} \supseteq E_k \cap \{y: \langle y - y^k, \nabla f(y^k) \rangle \leq 0\}$
  - **Stop** when the radius of the ellipsoid becomes $\leq \varepsilon R/M$

Invariant: If $f(y^*) \leq A$ then $y^* \in E_k$ for all $k$

**Proof:** Convexity of $f$ implies $\langle y^* - y^k, \nabla f(y^k) \rangle + f(y^k) \leq f(y^*) \leq A$
Since $f(y^k) > A$, $\langle y^* - y^k, \nabla f(y^k) \rangle < 0$

**Running Time:** $\text{poly}(m, t_f, t\nabla f, \log \frac{RM}{\varepsilon})$
Bounding $R$ and $M$?

$\inf_{y \in \mathbb{R}^m} \log p(e^y) - \sum \theta_i y_i$

$\sup_q \sum_{\tau \in \mathcal{F}} q(\tau) \log \frac{p(\tau)}{q(\tau)}$ \quad \Rightarrow M \leq m

$\square$ $q$ – prob. distribution over $\mathcal{F}$

$\square$ The expectation of $q$ is $\theta$

Bounding $R$?: As $\theta$ comes close to the boundary, $y^*$ must blow up. By how much?

Theorem [SinghV. '14, StraszakV. '17b]: If the unary complexity of all facets of the polytope is polynomial in $m$ then, $R \leq \text{poly}(m)$ – includes all combinatorial polytopes

Entropy interpretation seems important to obtain the bit complexity bounds
Summary and Challenges

- Some non-convex problems can be geodesically convex – find a metric!
- Geodesics and their study is a highly developed area in math and physics
- Working with geodesics may comes at additional costs

- Polynomial time algorithm for Brascamp-Lieb constant for rank >1?
- Entropy interpretation of Brascamp-Lieb for rank >1?
- Understanding functions that are geodesically convex?
- Develop more methods for geodesic convex optimization?
- Sampling from geodesically convex densities?