Lecture 3

Recap of last lecture.

(Nocedal & Wright Chap 12)

In Eucleadian setting, we discussed about contrained optimization problem.

$$\min_{x \in \Omega} f(x)$$

$\Omega \subset \mathbb{R}^d$, $f \in C^\infty(\mathbb{R}^n)$

Ω is also "smooth"

Given a point x^*, how to check if it is a local optima.

x^* local optimal $\Leftrightarrow$ $\exists$ a small neighborhood $U \subset \Omega$ around x^* s.t. $\forall\ x \in U$, $f(x^*) \leq f(x)$

Necessary condition for x^* to be locally optimal

If x^* is a local sol^n, then we have.

$\nabla f(x^*)^T d \geq 0 \quad \forall\ d$ s.t. d is a "tangent vector"

(is it a sufficient condition?)

What is a Tangent Vector

'd' is tangent to Ω at a point $x \in \Omega$. if there exist a sequence $\{z_k\}$ approaching x and a sequence $\{t_k\} > 0$ with $t_k \to 0$ such that

$$\lim_{k \to \infty} \frac{z_k - x}{t_k} = d.$$

Curve $\gamma : (a, b) \to \Omega$. , $0 \in (a, b)$
$\gamma(0) = x^*$, then.

$$\lim_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t} = \gamma'(0) = d.$$

$T_\Omega(x)$ is the collection of such tangent vectors
It is also called tangent cone.

But this definition is more "geometric".

In order to check if x^* is locally optimal,

you need to find all sequences $\{z_k\}$ s.t $z_k \in \Omega$

and $z_k \to x^*$ and then get their limiting directions

Difficult to do.

We want a more "algebraic" approach which can

be codified for computing.

Quest for "algebraic" necessary conditions of local optimality.

$$\min_{x \in \Omega} f(x) \qquad \text{where } \Omega \triangleq \{x \in \mathbb{R}^d \mid h(x) = 0\}$$

from first order analysis at x^* (using first-order Taylor approximation) we have the following:

① Construct a set $F_\Omega(x)$ of linearized feasible direction

$$F_\Omega(x) = \{d \mid \nabla h(x)^T d = 0\}$$

② If x^* is locally optimal sol^n, we get

$$\nabla f(x^*)^T d = 0 \quad \forall\ d \in F_\Omega(x^*)$$

The above gives a relationship bet^n $\nabla f(x^*)$ and $\nabla h(x^*)$

$$\nabla f(x^*) = \lambda \nabla h(x^*) \qquad \text{for some } \lambda \in \mathbb{R}$$

(what happens when we have $h_i(x) = 0$ for $i = 1 \ldots k$?)

In particular, we can write the first-order necessary conditions for x^* to be a local optimizer as

$$\| \operatorname{Proj}_{F_\Omega(x^*)} (\nabla f(x^*)) \| = 0$$

However, all this analysis works only when

$$F_\Omega(x) = T_\Omega(x),$$

One can always show that $T_\Omega(x) \subset F_\Omega(x)$

However, in order to show equivalence, we require some more conditions known as constraint qualification.

One such condition is Linear Independence Constraint Qualification (LICQ). It holds when the active constraint gradients $\{ \nabla h_i(x) \}$ are linearly independent.

Others also exist e.g. constant rank constraint qualification.

So, when LICQ holds at $x \in \Omega$.

$$T_\Omega(x) = \{ d \mid \langle \nabla h_i(x), d \rangle = 0 \}$$

and first order necessary condition of optimality is

$$\| \mathrm{Proj}_{T_\Omega(x^*)} (\nabla f(x^*)) \| = 0$$

- We analysed only first order necessary cond^n of local optimality
- It required that the set of tangent directions at x^* is a linear subspace of rank. ________ ?
assuming 'd' dimensions and k constraints

Our definition of smooth embedded submanifold is essentially that the above is true for all $x \in \mathcal{M}$ with a fixed rank 'k'.

for a manifold $\mathcal{M}$, the set of all tangent vectors at $x \in \mathcal{M}$ is called the "Tangent space" at x $(T_x\mathcal{M})$

For $S^{d-1} = \{ x \in \mathbb{R}^d \mid x^T x = 1 \}$

For $St(n,p) = \{ X \in \mathbb{R}^{n \times p} \mid X^T X = I_p \}$

Things we do in Euclidean Gradient Descent.

1. Current point $x_t \in \mathbb{R}^d$

2. Gradient $\nabla f(x_t)$

3. $x_{t+1} = x_t - \eta \nabla f(x_t)$

Manifold point of view.

1. Get point $x_t \in \mathcal{M}$ ✓

2. What is the meaning of gradient of a function in smooth non-linear space.

3. Operations like "+" or "−" are defined for linear spaces.
$\mathcal{M}$ is a non linear space

4. How to ensure $x_{t+1} \in \mathcal{M}$.

- In Euclidean setting $x_t - \eta \nabla f(x_t)$ may be interpreted as ?

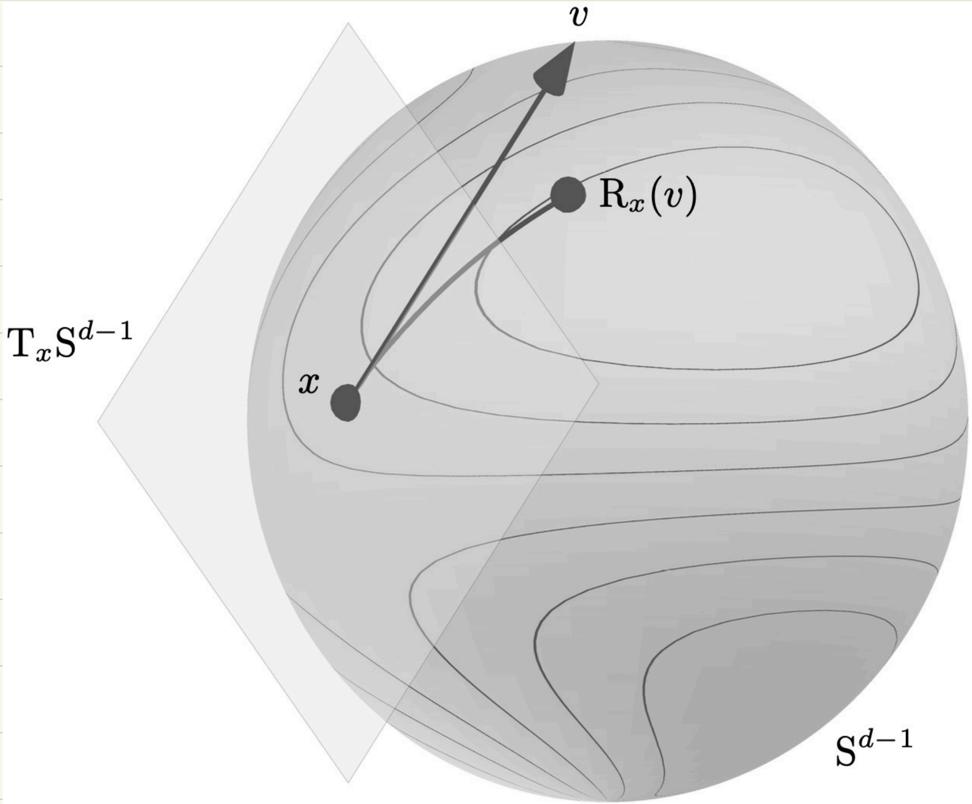
- How to ensure $x_{t+1} \in \mathcal{M}$?

Retraction at x : take a direction v at $x_t \in \mathcal{M}$ and give me the next point x_{t+1}

Are all directions 'v' allowed ?

Eg on $S^{d-1} = \{x \in \mathbb{R}^d \mid x^T x = 1\}$. $R_x : T_x S^{d-1} \to S^{d-1}$

$$v \to \frac{x+v}{\|x+v\|}$$


v
$R_x(v)$
T_xS^{d-1}
x
S^{d-1}

Notion of gradient

1. Euclidean setting: let $\mathcal{E}$ denote the Euclidean space

let $f : \mathcal{E} \to \mathbb{R}$, then.

$$Df(x) : \mathcal{E} \to \mathbb{R}, \quad Df(x)[v] = \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}$$

$$= \langle \nabla f(x), v \rangle \quad \forall v \in \mathcal{E}$$

Thus, we first chose the inner product space $\mathcal{E}$ and the gradient $\nabla f(x)$ definition depends on $\langle \cdot, \cdot \rangle$.

Now $T_x M$ is a linear space. Lets equip it with an inner product $\langle \, , \rangle_x$.

If this choice of inner product varies smoothly with x then we call it a Riemannian metric.

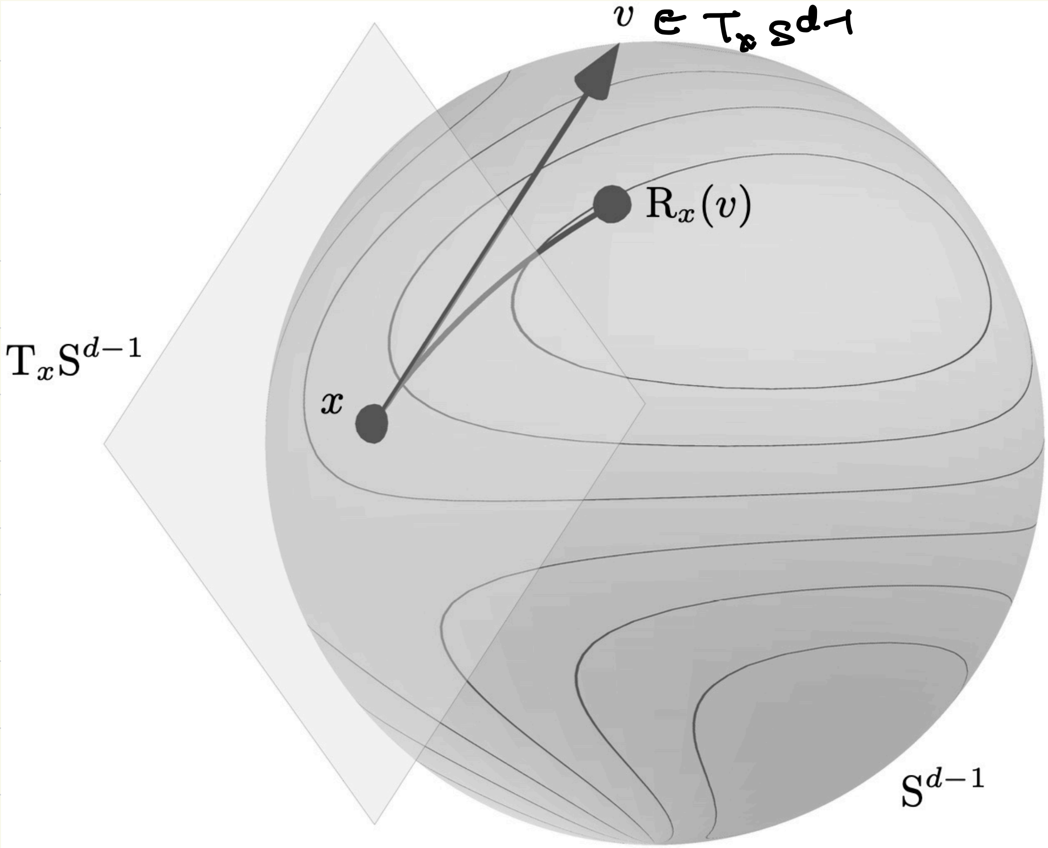
and M equipped with this metric is called Riemannian manifold.

Riemannian gradient: $\operatorname{grad} f(x)$ is the unique tangent vector at $x \in \mathcal{M}$ s.t. $\forall\, v \in T_x\mathcal{M}$

$$Df(x)[v] = \langle v, \operatorname{grad} f(x) \rangle_x$$

Lecture 4.

Recap from Lecture 3



$M \equiv S^{d-1}$
at $x \in M$.

Tangent space: $T_x M$

Retraction
$R_x : T_x M \to M$

Riemannian Metric: $\langle \cdot, \cdot \rangle_x$

M endowed with R. metric is called R. manifold.

One natural way to endow S^{d-1} with a metric.

Choose: $\langle \cdot, \cdot \rangle_x = \langle \cdot, \cdot \rangle \quad \forall x \in S^{d-1}$

Then we say that S^{d-1} is a Riemannian submanifold of $\mathbb{R}^d$

Riemannian gradient: $\operatorname{grad} f(x) \in T_x M$ s.t.

$$Df(x)[v] = \langle \underbrace{\operatorname{grad} f(x)}_{\text{(unique)}}, v \rangle_x \quad \forall v \in T_x M$$

This lecture.

- R. gradient for R. submanifolds
- Outline of R. gradient descent.
- Some demo
- Embedded submanifold of linear space (Chap. 3.2. N.B)
  - via local defining functions
  - via diffeomorphisms

Computing R.gradient for R.submanifolds.

let $\min_{x \in S^{d-1}} f(x)$

① Can we compute $\nabla f(x)$ (E. gradient) ?

Consider $\min_{x \in \mathbb{R}^d} \bar{f}(x)$

If $\bar{f}$ is smooth (i.e. $\bar{f} \in C^{\infty}(\mathbb{R}^d)$), we may expect that $\bar{f}$ restricted to domain S^{d-1} will also be smooth

$D\bar{f}(x)[v] = \langle \text{grad} \bar{f}(x), v \rangle$ where $x \in S^{d-1}$ and $v \in T_x S^{d-1}$

$Df(x)[v] = \langle \text{grad} f(x), v \rangle$

Since $D\bar{f}(x)[v] = Df(x)[v] \ \forall v \in T_x S^{d-1}$, $x \in S^{d-1}$

$\therefore \ \text{grad} f(x) \overset{?}{=} \text{grad} \bar{f}(x)$

For Riemannian submanifolds, the Riemannian gradient is an orthogonal projection of the "classical" gradient to the tangent spaces

$$\operatorname{grad} \bar{f}(x) = \operatorname{Proj}_{T_x S^{d-1}}\left(\operatorname{grad} \bar{f}(x)\right) + N_{T_x S^{d-1}}\left(\operatorname{grad} \bar{f}(x)\right)$$

$$\forall v \in T_x S^{d-1}, \langle N_{T_x S^{d-1}}(\operatorname{grad} \bar{f}(x)), v \rangle = 0$$

$$\therefore \quad \operatorname{grad} f(x) = \operatorname{Proj}_{T_x S^{d-1}}\left(\operatorname{grad} \bar{f}(x)\right)$$

Riemannian Gradient descent on S^{d-1}

1. Current point $x_t \in S^{d-1}$

2. Compute $\operatorname{grad} \bar{f}(x_t)$ where $\bar{f}: \mathbb{R}^d \to \mathbb{R}$ $(\bar{f}|_S = f)$ is a smooth extension of f.

3. Compute Riemannian gradient at x_t by Orthogonally projecting $\operatorname{grad} \bar{f}(x)$ onto $T_{x_t} S^{d-1}$

$$\operatorname{grad} f(x) = (I - xx^T) \operatorname{grad} \bar{f}(x)$$

4. Choose a step-size η.

5. Retraction: $x_{t+1} = R_{x_t}(v)$

where $v = -\eta \cdot \operatorname{grad} f(x)$

$$R_x(v) = \frac{x+v}{\|x+v\|} \quad \forall\, v \in T_x S^{d-1}$$

Demo.

Submanifolds of linear spaces

Consider a linear space $\mathcal{E}$ (e.g. $\mathbb{R}^d$)

What does it mean for a subset $M \subseteq \mathcal{E}$ to be "smooth"

We say a set M is smooth if locally around each point $x \in M$, we can say something like

"Around x, the set looks linear"

Let $M = \{ x \in \mathcal{E} \mid h(x) = 0 \}$ where $h : \mathcal{E} \to \mathbb{R}^k$

Then the truncated Taylor expansion of h is:

$$h(x+tv) = h(x) + t\,Dh(x)[v] + O(t^2)$$

$\therefore$ If $x \in M$ and $v \in \ker Dh(x)$, $h(x+tv) = O(t^2)$

$\therefore$ for small t, $h(x+tv) \simeq 0$ and subspace $\ker Dh(x)$ is the adequate linearization

Maybe, if h is smooth, then M is smooth.

Eg 1. $h(x) = x_1^2 + x_2^2 - 1$

Eg 2. $h(x) = x_1^2 - x_2^2$

eg. 3. $h(x) = x_1^2 - x_2^3$

eg 4: $h(x) = x_1^2 - x_2^4$

Embedded submanifold of $\mathcal{E}$. (Definition 1)

$\mathcal{E}$: linear space of dimension d.

M: non-empty subset of $\mathcal{E}$.

M is a (smooth) embedded submanifold of $\mathcal{E}$ of dimension 'n' if

1. $n = d$ and M is open in $\mathcal{E}$. (open submanifold)
2. $n = d - k$ for some (fixed) $k \geq 1$ and.
   - for each $x \in M$, there exists a neighborhood U of x in $\mathcal{E}$, and.
   - a smooth f^n $h : U \to \mathbb{R}^k$ such that

(a) $M \cap U = h^{-1}(0) \triangleq \{ y \in U : h(y) = 0 \}$

(b) $\operatorname{rank}(Dh(x)) = k$.

f^n h is called local defining f^n for M at x.

Embedded submanifolds of $\mathcal{E}$ (Definition 2)

Def: A diffeomorphism is a bijective map $F: U \to V$ where U, V are open sets such that F and F^{-1} are smooth.

Thm: Let $\mathcal{E}$ be a linear space of dim 'd'.

A subset M of $\mathcal{E}$ is an embedded submanifold of $\mathcal{E}$ of dimension $n = d-k$ iff.

for each $x \in M$, there exists

(a) a neighborhood U of x in $\mathcal{E}$, and

(b) an open subset V of $\mathbb{R}^d$, and

(c) a diffeomorphism $F: U \to V$ such tha

$F(M \cap U) = V \cap E$ where E is a linear subspace in $\mathbb{R}^d$ of dimension 'n'.