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. $\forall\ x \in U$, $\quad f(x^*) \leq f(x)$

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 directions

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

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 $i = 1 \dots k$?)

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

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

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 ?

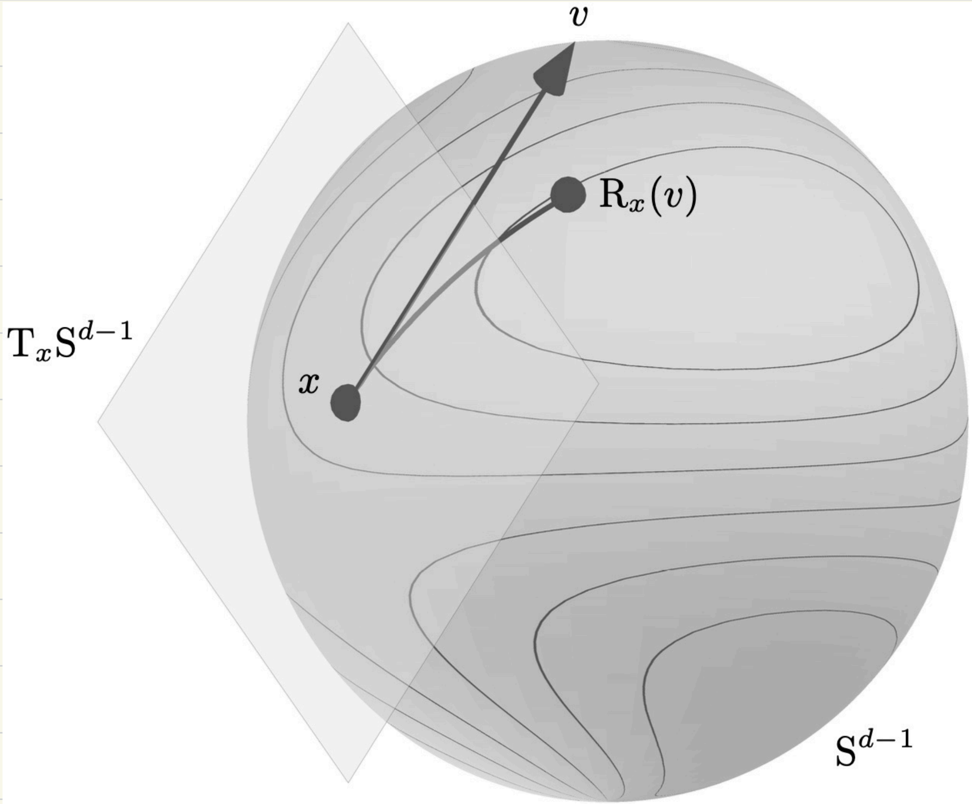
- How to ensure $x_{t+1} \in M$?

Retraction at x : take a direction v at $x_t \in M$ 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
$\mathrm{R}_x(v)$
$\mathrm{T}_x\mathrm{S}^{d-1}$
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: grad $f(x)$ is the unique tangent vector at $x \in \mathcal{M}$ s.t. $\forall$ $v \in T_x\mathcal{M}$

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

One natural way of endowing $S^{d-1} = \{x \in \mathbb{R}^d \mid x^T x = 1\}$ with a metric is to endow each tangent space $T_x S^{d-1}$ of $x \in S^{d-1}$ with the Euclidean inner product $\langle \cdot , \cdot \rangle$

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