Introduction

Sunday, January 5, 2025 9:09 AM

[illegible]

Algorithms are discrete processes.

input is discrete

output ———

steps are ———.

We will think of them as continuous.

Algorithm Discovery as a 3-step process:

- find the right space
- find the right path
- find the right discretization.

the ZEN of Algorithm Discovery.

$x \in \mathbb{R}^n \quad f: \mathbb{R}^n \to \mathbb{R}$

$\Omega \to \mathbb{R}$

$\Omega \subseteq \mathbb{R}^n.$

Two major problem classes:

<u>OPT</u>: given access to f, and $\epsilon > 0$,

find x s.t. $f(x) \leq \inf_{\Omega} f + \epsilon$

[illegible] distribution with

SAMPLING: Output x from distribution with density $\propto e^{-f}$.

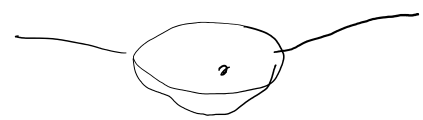
Both are intractable in general:

$$f(x) = \begin{cases} 0 & x = x^* \\ 1 & \text{o.w.} \end{cases}$$

even if f is continuous and bounded:

$$f : B^n(0,1) \longrightarrow [0,1]$$

$$f(x) = \min\{\varepsilon, \|x - x^*\|_2\}$$



n-dim ball of radius 1 centered at 0. $\Omega\left(\frac{1}{\varepsilon^n}\right)$ queries!

Similar examples even if f is differentiable with bounded derivative.

Convexity

f is convex if $\forall x, y \in \Omega$ $\forall \lambda \in [0,1]$

in Ω: $f((1-\lambda)x + \lambda y) \leq (1-\lambda) f(x) + \lambda f(y)$

$K \subseteq \mathbb{R}^n$ is convex if $\forall x, y \in K$ $\forall \lambda \in [0,1]$

$K \subseteq \mathbb{R}^n$ is convex if $\forall x, y \in K$ $\forall \lambda \in [0,1]$

$(1-\lambda)x + \lambda y \in K$.

Convex OPT

$\min_{x \in K} f(x)$ — convex.

Tractable!

Why?

Convex sets/functions are "separable".

Lemma. K is convex and closed $\Rightarrow$
$\subseteq \mathbb{R}^n$

$\forall y \notin K$, $\exists v \in \mathbb{R}^n$: $v^T y > v^T x$ $\forall x \in K$.

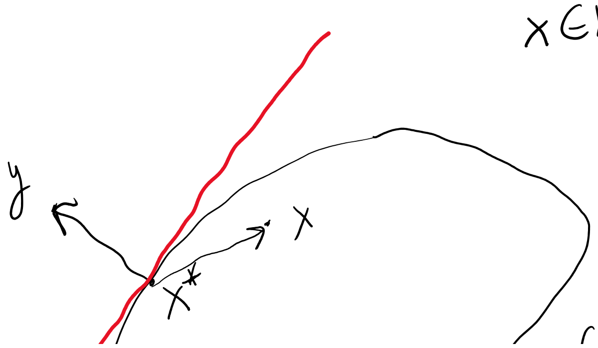
Pf.

$y \notin K$.

Let x^* be closest point to y in K,

(Exercise: x^* exists and is unique)

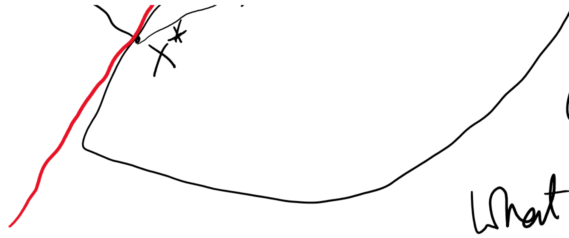
i.e. $x^* = \arg\min_{x \in K} \|x - y\|_2$



What is the candidate v?

$v = y - x^*$!

clearly $v^T(y - x^*) > 0$.



clearly $v^T(y-x^*) > 0$.

What about $v^T(x-x^*)$?

$$\forall t \in [0,1] \quad \| y - ((1-t)x^* + t x) \|^2 \geq \|y - x^*\|^2$$

$$\| y - x^* + t(x^* - x) \|^2 \geq \|y - x^*\|^2$$

$$t \|x^* - x\|^2 \geq \langle y - x^*, x - x^* \rangle$$

$$t \to 0 \qquad 0 \geq \langle y - x^*, x - x^* \rangle = v^T(x - x^*)$$

$$\therefore \quad v^T(y - x^*) > v^T(x - x^*)$$

$$\boxed{v^T y > v^T x}$$

This key property will enable "binary search" in high dimension!

Lema. $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then the following are equivalent: $\forall\, x, y \in \mathbb{R}^n$:

(1) $\forall t \in [0,1] \quad f((1-t)x + ty) \leq (1-t)f(x) + t f(y)$

(2) $f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$

(3) $\nabla^2 f(x) \succeq 0$

Pf. Fix x, y. Define $g: [0,1] \to \mathbb{R} \quad g(t) = f((1-t)x + ty)$.

(1) $\Rightarrow$ (2). $g(0) = f(x) \quad g(1) = f(y) \quad$ by (1), $g(t) \leq (1-t)g(0) + t g(1)$

$(1) \Rightarrow (2)$. $g(0) = f(x)$ $g(1) = f(y)$ by (1), $g(t) \le (1-t)g(0) + t\,g(1)$

$$\frac{g(t) - g(0)}{t} \le g(1) - g(0)$$

$$\lim_{t \to 0} \Rightarrow \quad g(1) - g(0) \geq g'(0)$$

What is $g'(t)$? $\langle \nabla f((1-t)x + ty), y - x \rangle$

So $f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$ (2).

Recall Taylor's thm.

Thm (4) $g: \mathbb{R} \to \mathbb{R}$ $(k+1)$-times differentiable.

$\forall x, y \in \mathbb{R}$

$$g(y) = g(x) + \sum_{i=1}^{k} g^{(i)}(x) \cdot \frac{(y-x)^i}{i!} + g^{(k+1)}(\xi) \cdot \frac{(y-x)^{k+1}}{(k+1)!}$$

for some $\xi \in [x, y]$.

Using $g(t) = f((1-t)x + ty)$

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(\xi)$$

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x)$$

$z \in [x, y]$

By (2) $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$.

$$\therefore \quad (y-x)^T \nabla^2 f(z)(y-x) \geq 0$$

Lim $\Rightarrow$ $\nabla^2 f(x) \succeq 0$. $(y = x + th, \ t \to 0)$.

$\lim_{y \to x} \Rightarrow \nabla^2 f(x) \geq 0.$ $(y = x + th, \ t \to 0).$

(3).

Let $h(t) = f((1-t)x + ty) - (1-t)f(x) - t f(y)$

$t^* = \arg\max_{t \in [0,1]} h(t).$ Goal: $\max h(t) \leq 0.$

if $t^* = \begin{cases} 0 \to h(0) = 0 \\ 1 \to h(1) = 0 \end{cases}$ ✓.

Assume $t^* \in (0,1).$

$$h(1) = h(t^*) + \underset{\downarrow}{(1-t^*) h'(t^*)} + \frac{1}{2}(1-t^*)^2 h''(\xi) \quad \text{for some } \xi \in [t^*, 1)$$

$$h(1) = h(t^*) + 0 + \underline{\geq 0} \quad \text{by (3)}$$

So $h(1) \geq h(t^*) \geq h(t)$

i.e. $0 \geq h(t)$ which is (1).

<u>Example.</u> "Logistic Regression".

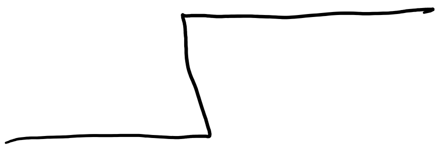
Input: pairs (x, y) $x \in \mathbb{R}^n$ $y \in \{-1, 1\}$

Goal: Find $\theta \in \mathbb{R}^n$ s.t. $\theta^T x > 0$ if $y = 1$

$\theta^T x < 0$ if $y = -1$

minimize $\text{Error} = \sum \mathbb{1}_{\{y(A^T x) > 0\}}$ ← not convex.

minimize $\text{Error} = \sum_{(x,y)} \mathbb{1}\{y(\theta^T x) > 0\}$ ← not convex.



$$\sigma(z) = \log(1 + e^z)$$

$$\text{Min } R(\theta) = \sum \sigma(y(\theta^T x)) + \|\theta\|$$

$$R'(\theta) = \sum \sigma'(y\theta^T x)\, y x$$

$$R''(\theta) = \sum \sigma''(y\theta^T x)\, y^2 x x^T \qquad y^2 = 1$$

$$\sigma'(z) = \frac{e^z}{1+e^z} = 1 - \frac{1}{1+e^z}$$

$$\sigma''(z) = \frac{e^z}{(1+e^z)^2} \geq 0 \implies R''(\theta) = \sum a_x x x^T, \quad a_x \geq 0$$

$$R''(\theta) \succeq 0$$

$\boxed{\text{R is convex!}}$

Course outline

Efficient Algorithms: polytime in input size.

necessary: "Good characterization": NP ∩ co-NP

necessary: "Good characterization": NP ∩ co-NP
short proofs of YES and NO answers to
decision problems (is $OPT < t$?)
search → decision

OPT, SAMP — frontier of P for opt, SAMP and related topics

Gradient Descent	Calculus
Elimination	Prob.
Reduction	Linear Algebra.
Geometrization	Algorithms
Sparsification	
Acceleration	
Discretization	