

Intro & Overview.

Monday, January 6, 2020 5:42 PM

What

Algorithms are discrete-time procedures to solve problems.

E.g. SORT

Does f have a SATISFYING assignment?

What should my next move be?

Translate "Vigyök boldog" to English

$$\boxed{91} \text{ optimization: } \min_{x \in \Omega} f(x)$$

very general. E.g. $\exists x: f(x) = 1$

$$\Leftrightarrow \min |f(x) - 1|$$

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

$$\min f(x)$$

takes an infinite # evaluations of f .

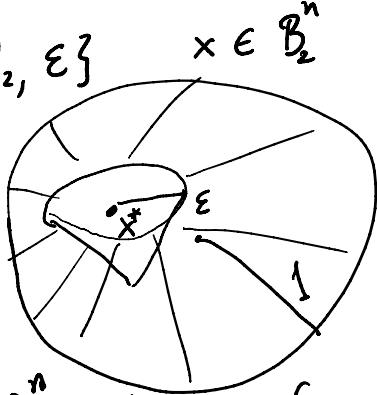
But what if f is Lipschitz, i.e. $\|\nabla f\| \leq 1$.

$L(f) = \sup_{x,y} \frac{|f(x) - f(y)|}{\|x - y\|}$ and with bounded domain.

$$f(x) = \min \left\{ \|x - x^*\|_2, \varepsilon \right\} \quad x \in \mathbb{B}_2^n$$

is 1 -Lipschitz.

$$\begin{aligned} \text{Vol}(\{x: \|x - x^*\|_2 \leq \varepsilon\}) \\ \leq \varepsilon^n \cdot \text{Vol}(\mathbb{B}_2^n) \end{aligned}$$



Final note: ε is the $o(1)^n$ call to f .

$$= \varepsilon \cdot \text{Var}(v_2)$$

Finding x takes $\mathcal{O}\left(\frac{1}{\varepsilon}\right)^n$ calls to f .

Exercise. Show that $\mathcal{O}\left(\frac{1}{\varepsilon}\right)^n$ calls suffice.

[P2] Sampling. $f: \mathbb{R}^n \rightarrow \mathbb{R}$, sample x with density $p(x) \propto e^{-f(x)}$.

$$\text{e.g. } f(x) = \frac{\|x\|^2}{2}, \quad p(x) = \begin{cases} 0 & x \in K \\ \infty & \text{o.w.} \end{cases}$$

OPT \rightarrow Sampling (binary search)

\therefore intractable!

Examples in practice: flow, matching, LP

regression, sampling Gaussian, convex body, logistic regression.

$$\frac{1}{n} \sum_{i=1}^n f(\alpha^T x_i \cdot y_i) + \lambda \| \alpha \|$$

Suppose we assume f is convex.

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

then all kinds of wonderful things happen!

① Any local minimum $\nabla f(x) = 0 \Rightarrow$ global min. $f(x^*) \leq f(x) \forall x$.

Why? Lemma. $f(y) \geq f(x) + \nabla f(x)^T (y-x)$

Pf. $g(\lambda) = f((1-\lambda)x + \lambda y) \quad g(0) = x \quad g(1) = y$

by convexity $g(\lambda) \leq (1-\lambda)g(0) + \lambda g(1)$.

$$g(1) \geq g(0) + \frac{g(1) - g(0)}{\lambda}$$

$$\lambda \rightarrow 0$$

$$g(1) \geq g(0) + g'(0)$$

$$g'(0) = \nabla f(x)^T (y-x)$$

$$f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

$$\text{e.g. } f(z) = \log(1 + e^z).$$

$$f(y) \geq f(x) + \nabla f(x)^T (y-x).$$

② Separation. K is closed, i.e. $x, y \in K \Rightarrow [x, y] \subseteq K$.

$$y \notin K \Rightarrow \exists a : a^T y > \max_{x \in K} a^T x.$$

This lets us do binary search!

Examples of convex functions.

Gradient Descent.

while $\|\nabla f(x)\| > \varepsilon$:

$$x \leftarrow x - h \nabla f(x)$$

Goal: Find x : $\|\nabla f(x)\| \leq \varepsilon$.

Q. how many iterations of GD? What h to use?

Why this goal: x is near-minimum in its neighborhood.

But even this is not true if ∇f can change quickly.

Assume $\|\nabla^2 f\|_{op} \leq L$ i.e. $\nabla^2 f$ is L -Lipschitz.

Thm. $\nabla^2 f$ is L -Lip, $\forall \varepsilon > 0$, starting at x^0 , we reach x with $\|\nabla f(x)\| \leq \varepsilon$ in at most

$$\frac{2L}{\varepsilon^2} (f(x_0) - f(x^*)) \text{ steps}$$

$$\text{Lem 1. } f\left(x - \frac{1}{L} \nabla f(x)\right) \leq f(x) - \frac{1}{2L} \|\nabla f(x)\|^2$$

Lemma 2. $\forall \lambda \in [0, 1]$

Pf. \exists [Taylor]. $\forall k+1$ times differentiable $g: \mathbb{R} \rightarrow \mathbb{R}$, $\forall x, y$

$$g(y) = g(x) + \sum_{i=1}^k \frac{(y-x)^i}{i!} g^{(i)}(x) + \frac{(y-x)^{k+1}}{(k+1)!} g^{(k+1)}(\zeta) \quad \zeta \in [x, y]$$

$$g(z) = g(0) + g'(0) + \frac{1}{2} g''(z) \quad z \in [0, 1].$$

$$g(t) = f((1-t)x + t y)$$

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

$y = x - \frac{1}{L} \nabla f(x)$ gives

$$\begin{aligned} f\left(x - \frac{1}{L} \nabla f(x)\right) &\leq f(x) - \frac{\| \nabla f(x) \|^2}{L} + \frac{1}{2} L \cdot \frac{\| \nabla f(x) \|^2}{L^2} \\ &\leq f(x) - \frac{1}{2L} \| \nabla f(x) \|^2. \end{aligned}$$

Start at $f(x^0)$, min inf $f(x^*)$. each step decrease is

$$\geq \frac{\varepsilon^2}{2L}.$$

 $\Rightarrow \# \text{ steps} \leq \frac{4L}{\varepsilon^2} (f(x^0) - f(x^*)).$

f convex

$$\text{Epi}(f) = \{(x, t) : f(x) \leq t\}$$

Exercise: $\text{Epi}(f)$ is convex.

$$f(x) = x^2$$

$$\min f(x) = \arg \min_t \text{Epi}(f).$$

\min convex functions \Leftrightarrow minimizing linear functions
over convex sets.

Next time: 6D for convex f.

Checking convexity.

Lemma: Equivalent:

- (1) f is convex
- (2) $\forall x, y \quad f(y) \geq f(x) + \nabla f(x)^T (y-x)$
- (3) $\forall x \quad \nabla^2 f(x) \succeq 0$.

Pf. (1) \Rightarrow (2) earlier.

(2) \Rightarrow (3) By Taylor $f(x+th) = f(x) + t \nabla f(x)^T h + \frac{t^2}{2} h^T \nabla^2 f(z) h$ $z \in [x, x+th]$

$$\Rightarrow h^T \nabla^2 f(z) h \geq 0 \quad \forall h$$

$$t \rightarrow 0 \Rightarrow \nabla^2 f(x) \succeq 0.$$

(3) \Rightarrow (1) $g(\lambda) = f(\lambda x + (1-\lambda)y) - \lambda f(x) - (1-\lambda)f(y)$

$$g'(\lambda) = \nabla f(\quad). (x-y) - f(x) + f(y)$$

$$g''(\lambda) = (x-y)^T \nabla^2 f(\quad)(x-y)$$

Let $\lambda^* = \arg\max g(\lambda)$. If $\lambda^* = 0 \text{ or } 1$, $\Rightarrow g(\lambda) \leq 0 \checkmark$

$$g(1) = g(\lambda^*) + g'(\lambda^*) + \frac{1}{2} g''(z)^T (\lambda^*-1)^2 \quad z \in [\lambda^*, 1]$$

$$0 = g(1) \geq g(\lambda^*) \geq g(\lambda). \quad \checkmark.$$