Lecture 26

Last class (Bonus class!)

Last time

- Heuristics to solve the subproblem.
- Descent
- Full method
- Guarantees.

Today

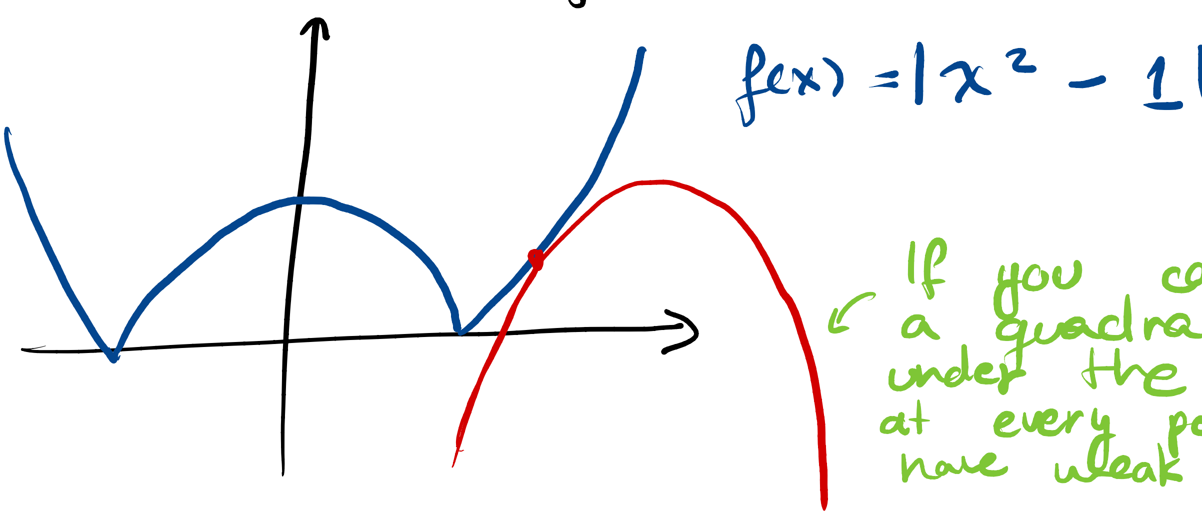
- Weakly convex functions
- Composite optimization
- A guarantee
- Closing remarks

Weakly convex functions

A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is called ρ-weakly convex if

$$x \mapsto f(x) + \frac{\rho}{2}\|x\|^2$$

is a convex function.



If you can fit a quadratic under the graph at every point, you have weak convexity.

Why is this an interesting class?

It gives a natural way to measure stationarity.

Def: A vector $\xi \in \mathbb{R}^d$ is a subgradient of a ρ-weakly convex function f at x ($\xi \in \partial f(x)$), if

$$\forall y \qquad f(y) \geq f(x) + \langle \xi, y-x \rangle - \frac{\rho}{2} \|x-y\|^2.$$

A point x is critical if $0 \in \partial f(x)$.

Proposition: Let $f: \mathbb{R}^d \to \mathbb{R}$ ρ-weakly convex, then for any $\lambda > 0$ with $\rho < \frac{1}{\lambda}$ the following are well-defined:

$$\text{prox}_{\lambda f}(x) = \underset{y}{\text{argmin}}\ f(y) + \frac{1}{2\lambda} \|y-x\|^2.$$

$$f_\lambda(x) = \min_y\ f(y) + \frac{1}{2\lambda} \|y-x\|^2.$$

Moreover, f_λ is continuously diff and if $\|\nabla f_\lambda(x)\| \leq \varepsilon$, then $x^+ = \text{prox}_{\lambda f}(x)$ satisfies:

i) $\|x - x^+\| \leq \lambda \varepsilon$

ii) $f(x^+) \leq f(x)$

iii) $\inf_{\xi \in \partial f(x^+)} \|\xi\| \leq \varepsilon$

Proof: Expanding

$$f(y) + \frac{1}{2\lambda}\|y-x\|^2 = f(y) + \frac{1}{2\lambda}\|y\|^2 + \langle x, y\rangle + \frac{1}{2\lambda}\|x\|^2$$

$$= \underbrace{f(y) + \frac{\rho}{2\lambda}\|y\|^2}_{\text{convex}} + \underbrace{\langle x, y\rangle + \frac{1}{2\lambda}\|x\|^2}_{\text{convex}}$$

$$+ \underbrace{\frac{1}{2}\left(\frac{1}{\lambda} - \rho\right)\|y\|^2}_{\text{strongly convex.}}$$

Since the function is strongly, everything is well-defined.

The fact that f_λ is C^1 follows from a similar reasoning from the HW, where you proved

$$\nabla f_\lambda(x) = \frac{1}{\lambda}(x - x^+).$$

Then, (i) follows trivially. Moreover, by definition of x^+:

$$f(x^+) \le f(x^+) + \frac{1}{2\lambda}\|x - x^+\|^2 \le f(x)$$

so (ii) follows.

This we will not prove.

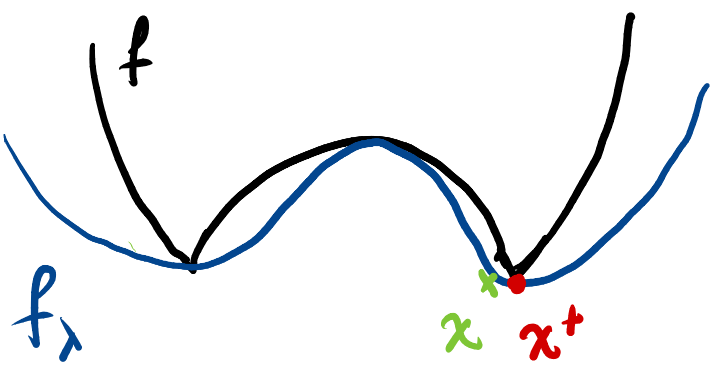
Finally, by the sum rule:

$$0 \in \partial f(x^+) + \frac{1}{\lambda}(x^+ - x)$$

$$\Rightarrow \|\nabla f_\lambda(x)\| = \frac{(x - x^+)}{\lambda} \in \partial f(x^+).$$

$\square$

Intuition



If we find x with $\|f_\lambda(x)\|$ small, then there is a close point that is almost stationary.

Composite optimization

Consider

L-smooth map

$$\min_x f(x) \quad \text{with} \quad f(x) = h \circ G(x)$$

β-Lips convex function

with $f: \mathbb{R}^m \to \mathbb{R}$ and $G: \mathbb{R}^d \to \mathbb{R}^m$.

This class of problems is weakly

convex and captures many data scientific tasks (phase retrieval, matrix completion, ...).

Let's consider two simple algorithms:

▷ Subgradient method

Update:

$$x_{k+1} \leftarrow x_k - \underbrace{\alpha_k}_{\text{stepsize}} \xi_k \quad \text{with } \xi_k \in \partial f(x_k)$$

One can show that $\partial f(x) = \nabla G(x)\, \partial h(x)$

▷ Gauss-Seidel method

Update:

$$x_{k+1} \leftarrow \operatorname{argmin} \Big\{ h\big(\overbrace{G(x_k) + \nabla G(x_k)(x - x_k)}^{\text{Linear approximation}}\big) + \underbrace{\frac{\beta}{2}}_{\text{stepsize.}} \|x - x_k\|^2 \Big\}$$

Note that the subgradient method applies to weakly convex problems, while Gauss Seidel applies to composite problems only.

One can show that subgradient descent achieves a rate of

$$\|\nabla f_\lambda(\bar{x}_k)\| = O\left(\frac{1}{k^{1/4}}\right) \leftarrow$$

Much slower than convex and smooth.

(Davis & Drusvyatskiy '18)

But local convergence might be much faster! Define $\operatorname{dist}(x, S) = \inf_{y \in S} \|x - y\|$

Theorem: Suppose that $f: \mathbb{R}^d \to \mathbb{R}$ is ρ-weakly convex, L-Lipschitz, and μ-sharp, i.e., let $S = \operatorname{argmin} f$,

$$\mu \operatorname{dist}(x, S) \leq f(x) - \min f.$$

If x_0 is such that $\operatorname{dist}(x_0, S) \leq \frac{1}{2}\frac{\mu}{\rho}$, then the iterates of subgradient descent with $\alpha_k = \dfrac{f(x_k) - \min f}{\|\xi_k\|^2}$ satisfy

$$\operatorname{dist}(x_{k+1}, S)^2 \leq \left(1 - \frac{\mu^2}{2L^2}\right) \operatorname{dist}(x_k, S)^2.$$

Proof: If x_0 lies in S there is nothing to prove as $\alpha_k = 0$. Let's show $\xi_k \neq 0$, assume it was zero, then $\exists\ \bar{x} \in S$

$$\mu \operatorname{dist}(x_0, S) = \mu \|x_0 - \bar{x}\| \leq f(x_0) - f(\bar{x}) \quad \leftarrow \text{sharpness}$$

Subgradient $\xi_0 = 0$ $\rightarrow$
$$< f(\bar{x}) + \frac{\rho}{2}\|x_0 - \bar{x}\|^2 - f(\bar{x})$$
$$= \rho \operatorname{dist}^2(x_0, S),$$

which contradicts $\operatorname{dist}(x_0, S) \leq \frac{1}{2}\frac{\mu}{\rho}$.

Then,

$$\|x_1 - \bar{x}\|^2$$
$$= \|x_0 - \alpha_0 \xi_0 - \bar{x}\|^2$$
$$= \|x_0 - \bar{x}\|^2 + 2\alpha_0 \langle \xi_0, \bar{x} - x_0 \rangle + \alpha_0^2 \|\xi_0\|^2$$
$$= \|x_0 - \bar{x}\|^2 + 2\frac{(f(x_0) - f^*)}{\|\xi_0\|^2}\langle \xi_0, \bar{x} - x_0 \rangle + \frac{(f(x_0) - f^*)^2}{\|\xi_0\|^2}$$
$$\leq \|x_0 - \bar{x}\|^2 + 2\frac{(f(x_0) - f^*)}{\|\xi_0\|^2}\left(f^* - f(x_0) + \frac{\rho}{2}\|x_0 - \bar{x}\|^2\right) + \frac{(f(x_0) - f^*)^2}{\|\xi_0\|^2}$$
$$= \|x_0 - \bar{x}\|^2 + \frac{(f(x_0) - f^*)}{\|\xi_0\|^2}\left(\rho\|x_0 - \bar{x}\|^2 - (f(x_0) - f^*)\right)$$
$$\leq \|x_0 - \bar{x}\|^2 + \frac{(f(x_0) - f^*)}{\|\xi_0\|^2}\left(\rho \underbrace{\|x_0 - \bar{x}\|^2}_{\leq \frac{1}{2}\frac{\mu}{\rho}} - \mu\|x_0 - \bar{x}\|\right)$$

$$\leq \|x_0 - \bar{x}\|^2 - \frac{\mu}{2} \frac{(f(x_0) - f^*)}{\|\xi_0\|^2} \|x_0 - \bar{x}\|$$

$$\leq \|x_0 - \bar{x}\|^2 - \frac{\mu^2}{2L} \|x_0 - \bar{x}\|.$$

μ-sharp and L-Lipschitz.

The proof follows by induction. $\square$

Closing remarks

We have build machinary to tackle

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

in a wide variety of settings.

- Optimality conditions
- First-order methods
  - Smooth opt
  - Nonsmooth opt
  - Stochastic / coordinate methods
  - Conjugate gradient
- Second order methods
  - Newton's
  - Quasi Newton
  - Gauss-Seidel
  - Trust region

THANK YOU!