ESE 415 Lecture 5 9/11/2024

Gradient Method

We will start discussing iterative optimization ALGORITHMS. Consider the following algorithm

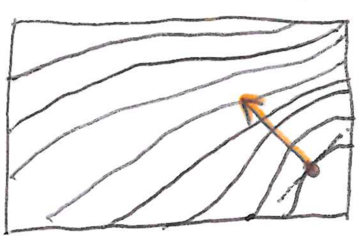
$$x^t = x^{t-1} - \gamma_t (\nabla f)(x^{t-1})$$

where $x^0 \in \mathbb{R}^n$ is the initialization and $\gamma_t > 0$ is called the step-size parameter.

This algorithm is called the gradient method (GM or steepest descent method.

Reasons for going in the direction of the gradient

- At any point $x \in \mathbb{R}^n$, the vector $-(\nabla f)(x)$ "points in the direction of the steepest descent of f" since $(D_v f)(x) = (\nabla f)(x)^T v$ is made most negative by chosing $v \in \mathbb{R}^n$ with $\|v\| = 1$ when
$$v = -\frac{(\nabla f)(x)}{\|(\nabla f)(x)\|}$$
- Recall level sets of $f(x)$:
$$L_f(c) := \{x \in \mathbb{R}^n : f(x) = c\}$$



"gradient is perpendicular to level sets" → most "efficient" step:

Theoretical Analysis of the Gradient Method:

For the analysis of the gradient method, we will consider functions with an additional property:

Def. A function f has a Lipschitz-continuous gradient with constant $L > 0$ when

$$\|(\nabla f)(x) - (\nabla f)(y)\| \leq L \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

We will denote this class of functions by $C_L^1(\mathbb{R}^n)$.

We will use the following auxiliary result later:

Lemma: For any $f \in C_L^1(\mathbb{R}^n)$, we have that

$$f(y) \leq f(x) + (\nabla f)(x)^T (y - x) + \frac{L}{2}\|y - x\|^2, \quad \forall x, y \in \mathbb{R}^n$$

Proof: HW3 problem!

We can now establish:

Theorem. Let $f \in C_L^1(\mathbb{R}^n)$ be a cost functional that has a finite minimum $f^* = f(x^*)$ attained at $x^* \in \mathbb{R}^n$. Then, for any step-size $\gamma \in (0, \frac{2}{L})$, the iterates generated by the gradient method satisfy

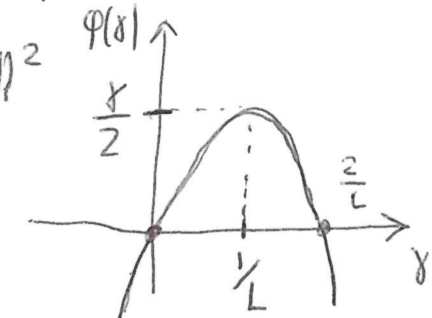
$$\lim_{t \to \infty} \|(\nabla f)(x^t)\| = 0.$$

<u>Proof</u>: Consider a single iteration

$$x^+ = x - \gamma (\nabla f)(x) \iff x^+ - x = -\gamma (\nabla f)(x)$$

Then for any $x \in \mathbb{R}^n$, we have that

$$\begin{aligned} f(x^+) &\leq f(x) + (\nabla f)(x)^T (x^+ - x) + \frac{L}{2} \|x^+ - x\|^2 \quad \text{Lemma} \\ &= f(x) + (\nabla f)(x)^T \left(-\gamma (\nabla f)(x)\right) + \frac{L \gamma^2}{2} \|(\nabla f)(x)\|^2 \\ &= f(x) - \gamma \|(\nabla f)(x)\|^2 + \frac{L\gamma^2}{2} \|(\nabla f)(x)\|^2 \\ &= f(x) - \underbrace{\gamma\left(1 - \frac{L\gamma}{2}\right)}_{\varphi(\gamma) := \gamma\left(1 - L\frac{\gamma}{2}\right)} \|(\nabla f)(x)\|^2 \end{aligned}$$



<u>Note</u>:

$\varphi(\gamma) > 0 \quad \forall \gamma \in (0, \frac{2}{L})$, max. at $\gamma^* = \frac{1}{L}$, $\varphi(\frac{1}{L}) = \frac{\gamma}{2}$.

Let $\gamma = \frac{1}{L}$ (without loss of generality), then

$$f(x^t) \leq f(x^{t-1}) - \frac{\gamma}{2} \|(\nabla f)(x^{t-1})\|^2$$

$$\Rightarrow \frac{\gamma}{2} \|(\nabla f)(x^{t-1})\|^2 \leq f(x^{t-1}) - f(x^t)$$

Telescoping sum by adding above inequality for $t = 0, 1, 2, \ldots$

$$\underbrace{\frac{\gamma}{2} \sum_{i=1}^{t} \|(\nabla f)(x^{i-1})\|^2}_{} \leq \left(f(x^0) - f(x^1)\right) + \left(f(x^1) - f(x^2)\right) + \ldots + f(x^{t-1}) - f(x^t)$$

series obtained by taking $\lim_{t \to \infty}$ exists, and therefore the summands converge to zero as $t \to \infty$:

$$= f(x^0) - \underbrace{f(x^t)}_{\geq f(x^*) \ldots \text{finite}} \leftarrow \text{finite}$$

$$\|(\nabla f)(x^t)\| \xrightarrow{t \to \infty} 0. \quad \blacksquare$$