Optimization Methods — *Linhui Xie*
Prof. Żak, Prof. Chong — *February 1, 2023*

8 Gradient Method

- Gradient algorithm:
$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f\left(x^{(k)}\right).$$

- Define $g^{(k)} := \nabla f\left(x^{(k)}\right)$ and set descent direction to $d^{(k)} = -g^{(k)}$.

- Step size α_k can be chosen in many different ways.

- For sufficiently small step size, the gradient algorithm has *descent* property.

  Define $\phi(\alpha) := f\left(x^{(k)} - \alpha g^{(k)}\right)$, then ϕ has Taylor expansion:
$$f\left(x^{(k)} - \alpha g^{(k)}\right) = f\left(x^{(k)}\right) - \alpha \left\|g^{(k)}\right\|^2 + o(\alpha)$$
  For α sufficiently small, we have
$$f\left(x^{(k)} - \alpha g^{(k)}\right) \le f\left(x^{(k)}\right)$$

► *PROPOSITION* Suppose $g^{(k)} = \nabla f\left(x^{(k)}\right) \neq \mathbf{0}$. There exists $\bar{\alpha} > 0$ such that for all $\alpha_k \in (0, \bar{\alpha})$, we have
$$f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$$

- Remark: if $g^{(k)} = \mathbf{0}$, the FONC holds.

  Short proof:

  By chain rule, we have
$$\phi'(0) = f\left(x^{(k)}\right) = -\left\|g^{(k)}\right\|^2 < 0.$$
  Gradient is negative thus function value is decreasing. Hence, there exists $\bar{\alpha} > 0$ such that for all $\alpha_k \in (0, \bar{\alpha})$, we have
$$\phi\left(\alpha_k\right) < \phi(0).$$
  Rewriting, we obtain
$$f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right).$$

- A variety of options exist for selecting α_k.

- If α_k too small, an increased number of iterations may be needed to get optimal solution x^*.

- If α_k too big, algorithm may lead to an oscillated track (zig-zag) around x^* (overshoot).

- General approach is to set a constant $\alpha_k = \alpha$ for all k.

- Steepest approach change α_k with each successive iteration.

8.1 Steepest descent algorithm

- Greedy scheme to pick for α_k.

$$\alpha_k = \underset{\alpha \geq 0}{\arg\min} f\left(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{g}^{(k)}\right).$$

- The steepest descent algorithm has the descent property.

► $\mathcal{PROPOSITION}$ 8.1 Let $\left\{\boldsymbol{x}^{(k)}\right\}$ be obtained by steepest descent method,

$$\left(\boldsymbol{x}^{(k+2)} - \boldsymbol{x}^{(k+1)}\right)^\top \left(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right) = 0.$$

Short Proof:

Based on definition,

$$\begin{aligned} \boldsymbol{x}^{(k+2)} &= \boldsymbol{x}^{(k+1)} - \alpha_{k+1}\boldsymbol{g}^{(k+1)}, \\ \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{g}^{(k)}. \end{aligned}$$

Let $\phi(\alpha) = f\left(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{g}^{(k)}\right) = f\left(\boldsymbol{x}^{(k+1)}\right)$. Since $\alpha_k = \arg\min \phi(\alpha)$, by FONC, we have

$$\phi'\left(\alpha_k\right) = 0.$$

Hence,

$$\phi'\left(\alpha_k\right) = \nabla f\left(\boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{g}^{(k)}\right)^\top \boldsymbol{g}^{(k)} = \nabla f\left(\boldsymbol{x}^{(k+1)}\right)^\top \boldsymbol{g}^{(k)} \overset{\boldsymbol{g}^{(k)} = \nabla f\left(\boldsymbol{x}^{(k)}\right)}{=\!=\!=} \boldsymbol{g}^{(k+1)^\top} \boldsymbol{g}^{(k)} = 0.$$

Therefore,

$$\left(\boldsymbol{x}^{(k+2)} - \boldsymbol{x}^{(k+1)}\right)^\top \left(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right) = \alpha_{k+1}\alpha_k \boldsymbol{g}^{(k+1)^\top} \boldsymbol{g}^{(k)} = 0.$$

- For a prescribed $\epsilon > 0$, terminate the iteration if one of the followings is met:
  - $\left\|\boldsymbol{g}^{(k)}\right\| < \epsilon$;
  - $\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right\| < \epsilon$;
  - $\left|f\left(\boldsymbol{x}^{(k+1)}\right) - f\left(\boldsymbol{x}^{(k)}\right)\right| < \epsilon$.
- More preferable choices using "relative change", because they are "scale-free".
  - $\left|f\left(\boldsymbol{x}^{(k+1)}\right) - f\left(\boldsymbol{x}^{(k)}\right)\right| / \left|f\left(\boldsymbol{x}^{(k)}\right)\right| < \epsilon$;
  - $\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right\| / \left\|\boldsymbol{x}^{(k)}\right\| < \epsilon$.

8.2 Analysis of optimization algorithms

- Globally convergent: starting from any initial point $\boldsymbol{x}^{(0)}$, an algorithm that generates sequence $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^*$, where $\boldsymbol{x}^*$ satisfying the FONC.
- Locally convergent: starting from an initial point $\boldsymbol{x}^{(0)}$ is sufficiently close to $\boldsymbol{x}^*$, an algorithm generates sequences converges to $\boldsymbol{x}^*$.
- Rate of convergence: how fast an algorithm converges.