

Fundamentals of Optimization

Subgradient method

Pham Quang Dung

dungpq@soict.hust.edu.vn

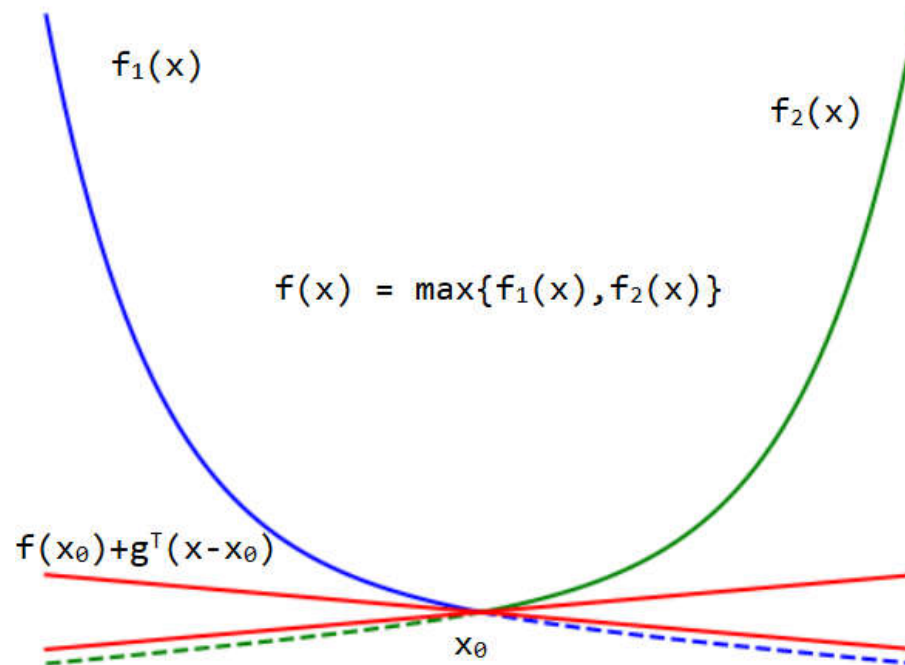
Department of Computer Science

Subgradient method

- For minimize nondifferentiable convex function
- Subgradient method is not a descent method: the function value can increase

Subgradient method

- Subgradient of f at x
 - Any vector g such that $f(x') \geq f(x) + g^T(x'-x)$
 - If f is differentiable, only possible choice is $g^{(k)} = \nabla f(x^{(k)})$,
→ the subgradient method reduces to the gradient method



Basic subgradient method

$$x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}$$

- $x^{(k)}$: is at the k^{th} iteration
- $g^{(k)}$: any subgradient of f at $x^{(k)}$
- $\alpha_k > 0$ is the k^{th} step size
- Note: subgradient is not a descent method, thus $f_{best}^{(k)} = \min\{f(x^{(1)}), f(x^{(2)}), \dots, f(x^{(k)})\}$

Convergence proof

- Notations: x^* is a minimizer of f
- Assumptions
 - Norm of the subgradients is bounded (with a constant G):
 $\|g^{(k)}\|_2 \leq G$ (this is the case if, for example, f satisfies the Lipschitz condition $|f(u) - f(v)| \leq G\|u-v\|_2$)
 - $\|x^{(1)} - x^*\|_2 \leq R$ (with a known constant R)
- We have $\|x^{(k+1)} - x^*\|_2^2 = \|x^{(k)} - \alpha_k g^{(k)} - x^*\|_2^2$
 $= \|x^{(k)} - x^*\|_2^2 - 2\alpha_k g^{(k)T}(x^{(k)} - x^*) + \alpha_k^2 \|g^{(k)}\|_2^2$
 $\leq \|x^{(k)} - x^*\|_2^2 - 2\alpha_k (f(x^{(k)}) - f(x^*)) + \alpha_k^2 \|g^{(k)}\|_2^2$ (due to the
fact that $f(x^*) \geq f(x^{(k)}) + g^{(k)T}(x^* - x^{(k)})$) (1)

Convergence proof

- Apply the inequality (1) recursively, we have

$$\|x^{(k+1)} - x^*\|_2^2 \leq \|x^{(1)} - x^*\|_2^2 - 2\sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^*) + \sum_{i=1}^k \alpha_i^2 \|g^{(i)}\|_2^2 \quad (\text{where } f^* = f(x^*))$$

$$\rightarrow 2\sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^*) \leq R^2 + \sum_{i=1}^k \alpha_i^2 \|g^{(i)}\|_2^2$$

$$\rightarrow R^2 + \sum_{i=1}^k \alpha_i^2 \|g^{(i)}\|_2^2 \geq 2\sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^*) \geq$$

$$2(\sum_{i=1}^k \alpha_i) \min_{i=1, \dots, k} (f(x^{(i)}) - f^*) = 2\sum_{i=1}^k \alpha_i (f_{best}^{(k)} - f^*)$$

$$\rightarrow f_{best}^{(k)} - f^* \leq \frac{R^2 + \sum_{i=1}^k \alpha_i^2 \|g^{(i)}\|_2^2}{2 \sum_{i=1}^k \alpha_i} \quad (2)$$

Convergence proof

- Different cases

- Constant step size $\alpha_k = \alpha$

$$\rightarrow f_{best}^{(k)} - f^* \leq \frac{R^2 + G^2 \alpha^2 k}{2\alpha k}$$

$$\rightarrow f_{best}^{(k)} - f^* \text{ converges to } G^2 \alpha / 2 \text{ when } k \rightarrow \infty$$

- Constant step length $\alpha_k = \gamma / \|g^{(k)}\|_2$

$$\rightarrow f_{best}^{(k)} - f^* \leq \frac{R^2 + \gamma^2 k}{2\gamma k / G}$$

$$\rightarrow f_{best}^{(k)} - f^* \text{ converges to } G\gamma / 2 \text{ when } k \rightarrow \infty$$

Convergence proof

- Different cases
 - Square summable but not summable

$$\|\alpha\|_2^2 = \sum_{i=1}^{\infty} \alpha_i^2 < \infty \text{ and } \sum_{i=1}^{\infty} \alpha_i = \infty$$

$$\rightarrow f_{best}^{(k)} - f^* \leq \frac{R^2 + G^2 \|\alpha\|_2^2}{2 \sum_{i=1}^k \alpha_i}$$

$$\rightarrow f_{best}^{(k)} - f^* \text{ converges to 0 as } k \rightarrow \infty$$

Example

$$\text{minimize } f(x) = \max_{i=1,\dots,m} (a_i^T x + b_i)$$

- Finding subgradient: given x , the index j for which

$$a_j^T x + b_j = \max_{i=1,\dots,m} (a_i^T x + b_i)$$

→ subgradient at x is $g = a_j$

Example

```
import numpy as np
def solve(A,b):
    f = lambda x: np.max(A.dot(x) + b)
    sg = lambda x: A[np.argmax(A.dot(x) + b)]
    x0 = [0,0,0,0]
    x = np.array(x0).T
    f_best = f(x)
    for i in range(100000):
        alpha = 2
        x = x - alpha*sg(x)
        if f_best > f(x):
            f_best = f(x)
    return f_best
```

Example

```
def main():  
    A = np.array([[1,-2,3,-5],[2,-2,1,1],[-3,2,-2,7]],dtype='double')  
    b = np.array([3,4,5]).T  
    rs = solve(A,b)  
    print('rs = ',rs)  
  
if __name__ == '__main__':  
    main()
```