

# On sliding gradient

Giulio Galvan

9 ottobre 2015

## 1 WCC analysis

**Step bound for generic descent directions.** From Lipschitz continuity we get:

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|x - y\|^2, \quad (1)$$

which choosing  $x = x_k$  and  $y = x_{k+1} = x_k + t_k d_k$  becomes:

$$f(x_{k+1}) - f(x_k) \leq t_k \nabla f(x_k)^T d_k + t_k^2 \frac{L}{2} \|d_k\|^2. \quad (2)$$

The previous inequality can be rewritten as:

$$f(x_k) - f(x_{k+1}) \geq -t_k \nabla f(x_k)^T d_k \cdot \left( 1 + t_k \frac{L}{2} \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2 \cos^2 \theta_k} \right) \quad (3)$$

where

$$\cos \theta_k = \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \|d_k\|} \quad (4)$$

In a backtracking setting, as defined in algorithm 1, we search for a value of  $t_k$  such that:

$$f(x_k) - f(x_{k+1}) \geq -\alpha t_k \nabla f(x_k)^T d_k. \quad (5)$$

When backtracking we have two possibilities: either  $t_k = s$  satisfy inequality (5) or not. In the latter case it must hold:

$$f(x_k) - f(x_k + \frac{t_k}{\beta} d_k) < -\alpha \frac{t_k}{\beta} \nabla f(x_k)^T d_k \quad (6)$$

Combining the latter with inequality 3 written for  $t_k = \frac{t_k}{\beta}$  yields:

$$-\alpha \frac{t_k}{\beta} \nabla f(x_k)^T d_k > -\frac{t_k}{\beta} \nabla f(x_k)^T d_k \cdot \left( 1 + \frac{t_k}{\beta} \frac{L}{2} \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2 \cos^2 \theta_k} \right), \quad (7)$$

which in turn, being  $\nabla f(x_k)^T d_k < 0$  since  $d_k$  is a descent direction, and  $t_k, \beta > 0$ , leads to:

$$t_k > \frac{2(\alpha - 1)\beta}{L} \frac{\|\nabla f(x_k)\|^2 \cos^2 \theta_k}{\nabla f(x_k)^T d_k} \quad (8)$$

$$= \frac{2(\alpha - 1)\beta}{L} \frac{\nabla f(x_k)^T d_k}{\|d_k\|^2} \quad (9)$$

If we impose

$$\delta \geq -\gamma \frac{\nabla f(x_k)^T d_k}{\|d_k\|^2} \quad (10)$$

where  $\gamma$  is some positive constant, we can use 8 and 10 in 5 and get:

$$f(x_k) - f(x_{k+1}) > \alpha \cdot \min \left( \gamma, \frac{2(1 - \alpha)\beta}{L} \right) \|\nabla f(x_k)\|^2 \cos^2 \theta_k. \quad (11)$$

Summing over  $k$ , if  $f$  is bounded below, say by  $f^*$ , and  $\theta_k$  bounded away from 90 degrees we get:

$$f(x_0) - f^* \geq \sum_{k=0}^N f(x_k) - f(x_{k+1}) = f(x_0) - f(x_N) > C \sum_{k=0}^N \|\nabla f(x_k)\|^2. \quad (12)$$

Hence we have convergence at the same rate as gradient descent.

We have made the following assumptions along the way:

- $f(\cdot)$  is bounded below
- $d_k$  is a descent direction bounded away from 90 degrees w.r.t  $\nabla f(x_k)$
- the initial guess of the backtracking algorithm  $\delta \geq -\gamma \frac{\nabla f(x_k)^T d_k}{\|d_k\|^2}$  for some positive constant  $\gamma$

---

**Algorithm 1:** Backtracking algorithm.

---

**Data:**

$\delta > 0$  initial step guess

$\alpha, \beta \in (0, 1)$

**1**  $t_k \leftarrow \delta$

**2 while**  $f(x_k) - f(x_{k+1}) < -\alpha t_k \nabla f(x_k)^T d_k$  **do**

**3**      $t_k \leftarrow t_k \beta$

**4 end**

**5 return**  $t_k$

---

**Convergence rate.** From inequality (12) we can immediately derive:

$$C(N+1) \min_{k \in (0, N)} \|\nabla f(x_k)\|^2 < f(x_0) - f^*, \quad (13)$$

hence

$$\min_{k \in (0, N)} \|\nabla f(x_k)\| < \frac{1}{\sqrt{(N+1)}} \left( \frac{f(x_0) - f^*}{C} \right)^{\frac{1}{2}}, \quad (14)$$

which proves that using descent directions different from the gradient still yields the same convergence rate of the gradient descent method  $\mathcal{O}(\frac{1}{\sqrt{N}})$ .

## 2 Sliding gradient framework

---

**Algorithm 2:** Backtracking algorithm.

---

**Data:**  
 $\mathbf{x}_0$  candidate solution  
1 **while** *stop criterion* **do**  
2    $i \leftarrow 0$   
3   **while**  $i < T$  *and gradient is degenerate* **do**  
4      $\mathbf{d}_{k,i} \leftarrow -\nabla f(x_{k-1})$  (or something smarter)  
5      $\mathbf{x}_{k,i} \leftarrow \mathbf{x}_{k,i} + \alpha_{k,i} \mathbf{d}_{k,i}$   
6   **end**  
7    $\mathbf{d}_k \leftarrow -\nabla f(x_k)$   
8    $\mathbf{x}_k \leftarrow \mathbf{x}_k + \alpha_k \mathbf{d}_k$   
9 **end**  
10 **return**  $t_k$

---