

# Optimization Theory

## Convergence of descent methods with backtracking (Armijo) linesearch

Anton Evgrafov

Department of Mathematical Sciences, NTNU [anton.evgrafov@math.ntnu.no](mailto:anton.evgrafov@math.ntnu.no)

**Read:** Section 3.1 in Nocedal and Wright, “Numerical optimization,” in particular Algorithm 3.1, p. 37.

Consider the following iteration:

$$x_{k+1} = x_k + \alpha_k p_k, \quad k = 0, 1, 2, \dots$$

where  $B_k = B_k^T$ ,

$$B_k p_k = -\nabla f(x_k),$$

and  $\alpha_k$  is selected using the backtracking (Armijo) linesearch with parameters  $c, \rho \in (0, 1)$ .

**Theorem 1.** *Suppose that*

1.  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable;
2. the set  $S := \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$  is bounded;
3. the matrices  $B_k$  are uniformly positive definite and bounded, that is  $\exists m > 0, M > 0 : m \leq \lambda_{\min}(B_k) \leq \lambda_{\max}(B_k) \leq M$ , where  $\lambda_{\min}$  and  $\lambda_{\max}$  are the smallest and the largest eigenvalues of  $B_k$ .

*Then the sequence  $\{x_k\}$  is bounded, and every its limit point  $\hat{x}$  is a stationary point for  $f$ .*

*Proof.* Owing to the sufficient descent condition in the linesearch procedure the sequence  $f(x_k)$ ,  $k = 0, 1, 2, \dots$  is non-increasing; thus  $x_k \in S$  for all  $k$ ; in particular it is bounded and therefore has at least one limit point. The set  $S$  is closed because  $f$  is continuous, and thus is compact owing to the assumption 2 and Heine–Borel theorem. Therefore, the function  $f$  attains its minimum value on  $S$  (Weierstrass theorem) and thus is bounded from below on  $S$ . As a result, the non-increasing sequence  $f(x_k)$  has a finite limit, and furthermore  $\lim_{k \rightarrow \infty} [f(x_{k+1}) - f(x_k)] = 0$ .

Owing to the sufficient descent condition it holds that

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq c\alpha_k \nabla f(x_k)^T p_k = -c\alpha_k \nabla f(x_k)^T B_k^{-1} \nabla f(x_k) \\ &\leq -c\alpha_k \lambda_{\max}(B_k^{-1}) \|\nabla f(x_k)\|^2 \\ &\leq -cM^{-1}\alpha_k \|\nabla f(x_k)\|^2 \leq 0. \end{aligned} \tag{1}$$

The sequence on the left converges to 0, meaning that the sequence on the right must also converge to zero. We will show that this implies that  $\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$ .

Suppose that this is not true; then, for some subsequence of indices  $k'$  and some  $\epsilon > 0$  we must have that  $\|\nabla f(x_{k'})\| \geq \epsilon$ . From (1) it then follows that  $\lim_{k' \rightarrow \infty} \alpha_{k'} = 0$ . In particular, it means that the step  $\alpha_{k'}/\rho$  was not acceptable to the linesearch procedure for all large  $k'$ , that is

$$f(x_{k'} + \alpha_{k'}\rho^{-1}p_{k'}) > f(x_{k'}) + c\alpha_{k'}\rho^{-1}\nabla f(x_{k'})^T p_{k'}. \quad (2)$$

The sequence of directions  $p_k = -B_k^{-1}\nabla f(x_k)$  is bounded. Indeed, by our assumption 3 the norms  $\|B_k^{-1}\| = \lambda_{\min}^{-1}(B_k) \leq m^{-1}$ . Furthermore, the continuous function  $x \mapsto \|\nabla f(x)\|$  attains its maximum over the compact set  $S$ , and thus  $\|\nabla f(x_k)\|$  is bounded by this maximum value, for all  $k$ . As a result, we may assume that for some subsequence of  $k'$ , say  $k''$ , it holds that  $\lim_{k'' \rightarrow \infty} x_{k''} = \hat{x}$  and  $\lim_{k'' \rightarrow \infty} p_{k''} = \hat{p}$ . Rearranging the terms in (2) we get

$$\begin{aligned} 0 &\leq \lim_{k'' \rightarrow \infty} \frac{f(x_{k''} + \alpha_{k''}\rho^{-1}p_{k''}) - f(x_{k''})}{\alpha_{k''}\rho^{-1}} - c\nabla f(x_{k''})^T p_{k''} \\ &= (1-c)\nabla f(\hat{x})^T \hat{p}, \end{aligned} \quad (3)$$

and therefore  $\nabla f(\hat{x})^T \hat{p} \geq 0$  as  $0 < c < 1$ . On the other hand,

$$\begin{aligned} \nabla f(\hat{x})^T \hat{p} &= \lim_{k'' \rightarrow \infty} \nabla f(x_{k''})^T p_{k''} = - \lim_{k'' \rightarrow \infty} \nabla f(x_{k''})^T B_{k''}^{-1} \nabla f(x_{k''}) \\ &\leq -M^{-1}\epsilon^2 < 0. \end{aligned} \quad (4)$$

However, equations (3) and (4) contradict each other. This must mean that our assumption that  $\|\nabla f(x_{k'})\| \geq \epsilon$  over some subsequence  $k'$  is wrong and in fact

$$\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0. \quad (5)$$

Finally, let  $\hat{x}$  be an arbitrary limit point of  $\{x_k\}$ , that is,  $\hat{x} = \lim_{k'' \rightarrow \infty} x_{k''}$  for some subsequence  $k''$ . Owing to the continuity of the function  $x \mapsto \|\nabla f(x)\|$  (assumption 1) and (5) it holds that  $\|\nabla f(\hat{x})\| = 0$ , as we claimed.  $\square$

**Exercise:** Using this theorem, show that the steepest descent algorithm with Armijo (backtracking) linesearch converges to the minimum of Rosenbrock function from any starting point.