Proof of convergence of steepest descent, without using algorithmic maps

Ruszczynski contains several proofs under different conditions. We sketch one.

THM Assume f is continuously differentiable and its gradient is Lipschitz continuous with constant M, so $\|\nabla f(x)-\nabla f(y)\| \le M\|x-y\|$ for any x, y.
Assume the set $X_0 := \{x : f(x) \le f(x^0)\}$ is bounded, where x_0 is the initial iterate.
Assume the steplength τ_k satisfies the two-slope Armijo test, so

$$f(x^k) + \alpha_1 \tau_k d^T \nabla f(x^k) \le f(x^k + \tau_k d) \le f(x^k) + \alpha_2 \tau_k d^T \nabla f(x^k)$$

for some α_1, α_2 satisfying $0 < \alpha_2 < \alpha_1 < 1$.

Then the sequence of iterates $\{x^k\}$ is bounded and every accumulation point x^* satisfies $\nabla f(x^*) = 0$

Sketch of proof

Note that $d = -\nabla f(x^k)$ with the method of steepest descent.

So we have

$$(*) \quad f(x^k) - \alpha_1 \tau_k \|\nabla f(x^k)\|^2 \le f(\overbrace{x^k - \tau_k \nabla f(x^k)}^{x^{k+1}}) \le f(x^k) - \alpha_2 \tau_k \|\nabla f(x^k)\|^2$$

Right hand inequality immediately gives $x^{k+1} \in X^0$, and $f(x^{k+1}) \le f(x^k)$.
Since X^0 is bounded and f is continuous, $f(\cdot)$ is bounded below on X^0.
So $\tau_k \|\nabla f(x^k)\|^2 \to 0$

It can be shown using the left hand of $(*)$ that $\tau_k \ge \frac{1-\alpha_1}{M}$
(the function doesn't turn up so quickly because of Lipschitz, so τ_k is bounded below). Thus $\|\nabla f(x^k)\| \to 0$ as $k \to \infty$ $\boxed{\text{M}}$