Data: $x_0 \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $0 < c_1 < 1$, $0 < \rho < 1$
Result: Estimate of local minimum x_{k+1} of f
$\beta_0 = 1$;
for $k = 0, \ldots,$ *until any stopping condition is fulfiled* **do**
  $p_k = -\nabla f(x_k)$;
  $\alpha_k = \text{backtrack}(x_k, p_k, \beta_k, c_1, \rho)$;
  $x_{k+1} = x_k + \alpha_k p_k$;
  $\beta_{k+1} = \frac{\alpha_k}{\rho}$;

Algorithm 3: Steepest Descent

descent direction (12), $p_k = -\nabla f(x_k)$. By Zoutendijk's Lemma, Lemma 1, this choice ensures convergence to the optimum, assuming α_k fulfil the Wolfe conditions, Definition 3. We then use the backtracking linesearch with initial length β_k to find α_k and then perform the step. This procedure ensures that the sufficient decrease condition is fulfiled.

The choice of β_k is important for two reasons. On the one hand, if β_k is too long, then we will spend a lot of function evaluations on shrinking it. In cases where the function is expensive to compute, this would slow down the algorithm significantly. On the other hand, if step sizes are chosen too small, they might not fulfil the sufficient decrease condition and we risk slow or premature convergence.

We solve both problems by picking $\beta_{k+1} = \alpha_k/\rho$. The parameter $0 < \rho < 1$ is used by the backtracking linesearch to decrease the steplength by factor ρ, when the sufficient decrease conditions are not fulfilled. Thus, if the new step length is too long, we only perform one additional step of backtracking. But if at some point our choice of steplength becomes too small, the backtracking linesearch will accept the proposed longer steplength and thus, over several iterations the step size will increase.

Unfortunately, it is difficult to proof better convergence properties than what we have done in Lemma 1. It is fair to say, that on most functions, Steepest Descent will eventually converge linearly as measured on $\|\nabla f(x_k)\|$. Most theoretical results however rely on exact linesearches, something we will seldom have in practice. However, we should not undersell the result in Lemma 1: Under the condition that the Wolfe condition holds, this theorem proofs *global* convergence, that means no matter the starting point Steepest Descent will always converge to a point fulfilling the first order sufficient conditions, even though it might sometimes take a long time.

Newton's Algorithm is probably the most well known fast optimisation algorithm. As we will show in Section 3, this algorithm has *local* quadratic convergence towards a strict local optimum. When the algorithm is started sufficiently close to the optimum it will converge at a quadratic rate. Most surprisingly, this holds even when choosing $\alpha_k = 1$.

The algorithm is based on choosing the step that minimises the quadratic