(1) If $\lambda(u_k) > \eta$, then
$$J(u_{k+1}) - J(u_k) \leq -\gamma.$$

(2) If $\lambda(u_k) \leq \eta$, then the backtracking line search selects $t = 1$ and we have
$$2\lambda(u_{k+1}) \leq (2\lambda(u_k))^2.$$

As a consequence, for all $\ell \geq k$, we have
$$J(u_\ell) - p^* \leq \lambda(u_\ell)^2 \leq \left(\frac{1}{2}\right)^{2^{\ell-k+1}}.$$

In the end, accuracy $\epsilon > 0$ is achieved in at most
$$\frac{20 - 8\alpha}{\alpha\beta(1-2\alpha)^2}(J(u_0) - p^*) + \log_2 \log_2(1/\epsilon)$$
iterations, where α and β are the constants involved in the line search. This bound is obviously independent of the chosen coordinate system.

Contrary to intuition, the descent direction $d_k = -\nabla J_{u_k}$ given by the opposite of the gradient is *not* always optimal. In the next section we will see how a better direction can be picked; this is the method of *conjugate gradients*.

49.10 Conjugate Gradient Methods for Unconstrained Problems

The conjugate gradient method due to Hestenes and Stiefel (1952) is a gradient descent method that applies to an elliptic quadratic functional $J\colon \mathbb{R}^n \to \mathbb{R}$ given by
$$J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle,$$
where A is an $n \times n$ symmetric positive definite matrix. Although it is presented as an iterative method, it terminates in at most n steps.

As usual, the conjugate gradient method starts with some arbitrary initial vector u_0 and proceeds through a sequence of iteration steps generating (better and better) approximations u_k of the optimal vector u minimizing J. During an iteration step, two vectors need to be determined:

(1) The descent direction d_k.

(2) The next approximation u_{k+1}. To find u_{k+1}, we need to find the stepsize $\rho_k > 0$ and then
$$u_{k+1} = u_k - \rho_k d_k.$$
Typically, ρ_k is found by performing a line search along the direction d_k, namely we find ρ_k as the real number such that the function $\rho \mapsto J(u_k - \rho d_k)$ is minimized.