We saw in Proposition 49.13 that during execution of the gradient method with optimal stepsize parameter that any two consecutive descent directions are orthogonal. The new twist with the conjugate gradient method is that given $u_0, u_1, \ldots, u_k$, the next approximation u_{k+1} is obtained as the solution of the problem which consists in minimizing J over the affine subspace $u_k + \mathcal{G}_k$, where $\mathcal{G}_k$ is the subspace of $\mathbb{R}^n$ spanned by the gradients

$$\nabla J_{u_0}, \nabla J_{u_1}, \ldots, \nabla J_{u_k}.$$

We may assume that $\nabla J_{u_\ell} \neq 0$ for $\ell = 0, \ldots, k$, since the method terminates as soon as $\nabla J_{u_k} = 0$. A priori the subspace $\mathcal{G}_k$ has dimension $\leq k + 1$, but we will see that in fact it has dimension $k + 1$. Then we have

$$u_k + \mathcal{G}_k = \left\{ u_k + \sum_{i=0}^{k} \alpha_i \nabla J_{u_i} \;\middle|\; \alpha_i \in \mathbb{R},\ 0 \leq i \leq k \right\},$$

and our minimization problem is to find u_{k+1} such that

$$u_{k+1} \in u_k + \mathcal{G}_k \quad \text{and} \quad J(u_{k+1}) = \inf_{v \in u_k + \mathcal{G}_k} J(v).$$

In the gradient method with optimal stepsize parameter the descent direction d_k is proportional to the gradient ∇J_{u_k}, but in the conjugate gradient method, d_k is equal to ∇J_{u_k} corrected by some multiple of d_{k-1}.

The conjugate gradient method is superior to the gradient method with optimal stepsize parameter for the following reasons proved correct later:

(a) The gradients ∇J_{u_i} and ∇J_{u_j} are orthogonal for all i, j with $0 \leq i \neq j \leq k$. This implies that if $\nabla J_{u_i} \neq 0$ for $i = 0, \ldots, k$, then the vectors ∇J_{u_i} are linearly independent, so the method stops in at most n steps.

(b) If we write $\Delta_\ell = u_{\ell+1} - u_\ell = -\rho_\ell d_\ell$, the second remarkable fact about the conjugate gradient method is that the vectors Δ_ℓ satisfy the following conditions:

$$\langle A\Delta_\ell, \Delta_i \rangle = 0 \quad 0 \leq i < \ell \leq k.$$

The vectors Δ_ℓ and Δ_i are said to be *conjugate* with respect to the matrix A (or *A-conjugate*). As a consequence, if $\Delta_\ell \neq 0$ for $\ell = 0, \ldots, k$, then the vectors Δ_ℓ are linearly independent.

(c) There is a simple formula to compute d_{k+1} from d_k, and to compute ρ_k.

We now prove the above facts. We begin with (a).

Proposition 49.15. *Assume that $\nabla J_{u_i} \neq 0$ for $i = 0, \ldots, k$. Then the minimization problem, find u_{k+1} such that*

$$u_{k+1} \in u_k + \mathcal{G}_k \quad \textit{and} \quad J(u_{k+1}) = \inf_{v \in u_k + \mathcal{G}_k} J(v),$$

has a unique solution, and the gradients ∇J_{u_i} and ∇J_{u_j} are orthogonal for all i, j with $0 \leq i \neq j \leq k + 1$.