Remarks:

(1) Since A is symmetric positive definite, the bilinear map $(u, v) \mapsto \langle Au, v\rangle$ is an inner product $\langle -, -\rangle_A$ on $\mathbb{R}^n$. Consequently, two vectors u, v are *conjugate* with respect to the matrix A (or A-*conjugate*), which means that $\langle Au, v\rangle = 0$, iff u and v are orthogonal with respect to the inner product $\langle -, -\rangle_A$.

(2) By picking the descent direction to be $-\nabla J_{u_k}$, the gradient descent method with optimal stepsize parameter treats the level sets $\{u \mid J(u) = J(u_k)\}$ as if they were spheres. The conjugate gradient method is more subtle, and takes the "geometry" of the level set $\{u \mid J(u) = J(u_k)\}$ into account, through the notion of conjugate directions.

(3) The notion of conjugate direction has its origins in the theory of projective conics and quadrics where A is a 2×2 or a 3×3 matrix and where u and v are conjugate iff $u^\top Av = 0$.

(4) The terminology conjugate gradient is somewhat misleading. It is not the gradients who are conjugate directions, but the descent directions.

By definition of the vectors $\Delta_\ell = u_{\ell+1} - u_\ell$, we can write

$$\Delta_\ell = \sum_{i=0}^{\ell} \delta_i^\ell \, \nabla J_{u_i}, \quad 0 \leq \ell \leq k. \tag{$*_2$}$$

In matrix form, we can write

$$\begin{pmatrix} \Delta_0 & \Delta_1 & \cdots & \Delta_k \end{pmatrix} = \begin{pmatrix} \nabla J_{u_0} & \nabla J_{u_1} & \cdots & \nabla J_{u_k} \end{pmatrix} \begin{pmatrix} \delta_0^0 & \delta_0^1 & \cdots & \delta_0^{k-1} & \delta_0^k \\ 0 & \delta_1^1 & \cdots & \delta_1^{k-1} & \delta_1^k \\ 0 & 0 & \cdots & \delta_2^{k-1} & \delta_2^k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \delta_k^k \end{pmatrix},$$

which implies that $\delta_\ell^\ell \neq 0$ for $\ell = 0, \ldots, k$.

In view of the above fact, since Δ_ℓ and d_ℓ are collinear, it is convenient to write the descent direction d_ℓ as

$$d_\ell = \sum_{i=0}^{\ell-1} \lambda_i^\ell \, \nabla J_{u_i} + \nabla J_{u_\ell}, \quad 0 \leq \ell \leq k. \tag{$*_3$}$$

Our next goal is to compute u_{k+1}, assuming that the coefficients λ_i^k are known for $i = 0, \ldots, k$, and then to find simple formulae for the λ_i^k.

The problem reduces to finding ρ_k such that

$$J(u_k - \rho_k d_k) = \inf_{\rho \in \mathbb{R}} J(u_k - \rho d_k),$$