imply by induction that the subspace $\mathcal{G}_k$ is spanned by $(r_0, r_1, \ldots, r_k)$ and $(d_0, d_1, \ldots, d_k)$ is the subspace spanned by

$$(r_0, Ar_0, A^2 r_0, \ldots, A^k r_0).$$

Such a subspace is called a *Krylov subspace*.

If we define the *error* e_k as $e_k = u_k - u$, then $e_0 = u_0 - u$ and $Ae_0 = Au_0 - Au = Au_0 - b = d_0 = r_0$, and then because

$$u_{k+1} = u_k - \rho_k d_k$$

we see that

$$e_{k+1} = e_k - \rho_k d_k.$$

Since d_k belongs to the subspace spanned by $(r_0, Ar_0, A^2 r_0, \ldots, A^k r_0)$ and $r_0 = Ae_0$, we see that d_k belongs to the subspace spanned by $(Ae_0, A^2 e_0, A^3 e_0, \ldots, A^{k+1} e_0)$, and then by induction we see that e_{k+1} belongs to the subspace spanned by $(e_0, Ae_0, A^2 e_0, A^3 e_0, \ldots, A^{k+1} e_0)$. This means that there is a polynomial P_k of degree $\leq k$ such that $P_k(0) = 1$ and

$$e_k = P_k(A) e_0.$$

This is an important fact because it allows for an analysis of the convergence of the conjugate gradient method; see Trefethen and Bau [176] (Lecture 38). For this, since A is symmetric positive definite, we know that $\langle u, v \rangle_A = \langle Av, u \rangle$ is an inner product on $\mathbb{R}^n$ whose associated norm is denoted by $\|v\|_A$. Then observe that if $e(v) = v - u$, then

$$\begin{aligned}
\|e(v)\|_A^2 &= \langle Av - Au, v - u \rangle \\
&= \langle Av, v \rangle - 2\langle Au, v \rangle + \langle Au, u \rangle \\
&= \langle Av, v \rangle - 2\langle b, v \rangle + \langle b, u \rangle \\
&= 2J(v) + \langle b, u \rangle.
\end{aligned}$$

It follows that $v = u_k$ minimizes $\|e(v)\|_A$ on $u_{k-1} + \mathcal{G}_{k-1}$ since u_k minimizes J on $u_{k-1} + \mathcal{G}_{k-1}$. Since $e_k = P_k(A) e_0$ for some polynomial P_k of degree $\leq k$ such that $P_k(0) = 1$, if we let $\mathcal{P}_k$ be the set of polynomials $P(t)$ of degree $\leq k$ such that $P(0) = 1$, then we have

$$\|e_k\|_A = \inf_{P \in \mathcal{P}_k} \|P(A) e_0\|_A.$$

Since A is a symmetric positive definite matrix it has real positive eigenvalues $\lambda_1, \ldots, \lambda_n$ and there is an orthonormal basis of eigenvectors $h_1, \ldots, h_n$ so that if we write $e_0 = \sum_{j=1}^n a_j h_j$. then we have

$$\|e_0\|_A^2 = \langle Ae_0, e_0 \rangle = \left\langle \sum_{i=1}^n a_i \lambda_i h_i, \sum_{j=1}^n a_j h_j \right\rangle = \sum_{j=1}^n a_j^2 \lambda_j$$

and

$$\|P(A) e_0\|_A^2 = \langle AP(A) e_0, P(A) e_0 \rangle = \left\langle \sum_{i=1}^n a_i \lambda_i P(\lambda_i) h_i, \sum_{j=1}^n a_j P(\lambda_j) h_j \right\rangle = \sum_{j=1}^n a_j^2 \lambda_j (P(\lambda_j))^2.$$