Theorem 1. Assume that $A$ is an $n \times n$ real matrix which satisfies
\begin{align*}
\alpha \|x\|^2 \leq \langle Ax, x \rangle, & \quad \text{for all } x \in \mathbb{R}^n, \\
\langle Ax, y \rangle \leq \beta \|x\| \|y\|, & \quad \text{for all } x, y \in \mathbb{R}^n.
\end{align*}
Suppose that $e_i = x - x_i$ where $x_i$ is the $i$’th iterate in the GMRES algorithm (with starting iterate $x_0$) and $x$ is the solution of $Ax = b$. Then
\[ \|Ae_i\| \leq \rho^{i/2} \|Ae_0\| \]
for
\[ \rho = 1 - \frac{\alpha^2}{\beta^2}. \]

Proof. We consider the inner product
\[ \langle x, y \rangle = \langle Ax, Ay \rangle \quad \text{for all } x, y \in \mathbb{R}^n. \]
We shall denote its corresponding norm by $\| \cdot \|_* = (\langle \langle x, y \rangle \rangle)^{1/2}$. It follow from (15.1) that
\begin{align*}
\langle Ax, x \rangle & = \langle A^2x, Ax \rangle \geq \alpha \|Ax\|^2 = \alpha\|x\|_*, \\
\langle Ax, y \rangle & \leq \beta \|Ax\| \|Ay\| = \beta \|x\|_* \|y\|_*.
\end{align*}
We consider the Richardson method
\[ \tilde{x}_{i+1} = \tilde{x}_i + \tau (b - A\tilde{x}_i) \]
using the same initial vector as in the GMRES iteration, i.e., $\tilde{x}_0 = x_0$. Let $\tilde{e}_i = x - \tilde{x}_i$.

Then (15.2) and (15.3) and the proposition of last class implies that
\[ \|\tilde{e}_{i+1}\|_* \leq \sqrt{\rho} \|\tilde{e}_i\|_* . \]
Repetitively applying the above inequality gives
\[ \|\tilde{e}_i\|_* \leq \rho^{i/2} \|e_0\|_* . \]
Now,
\[ \tilde{e}_i = (I - \tau A)^i e_0 \]
so
\[ A(\tilde{e}_i) = (I - \tau A)^i r_0 = r_0 - A\zeta \]
for some \( \zeta \in K_i \). By the minimization property of GMRES,
\[
\|Ae_i\| = \|r(\theta)\| = \min_{\zeta \in K_i(A)} \|r(\zeta)\|
= \min_{\zeta \in K_i(A)} \|r_0 - A\zeta\| \leq \|A\tilde{e}_i\|
= \|\tilde{e}_i\| \leq \rho^{i/2} \|e_0\| = \rho^{i/2} \|Ae_0\|.
\]
This completes the proof of the theorem. \(\square\)

An alternative to applying GMRES is to apply CG to the normal equations. There are some who argue against this approach as it generally squares the condition number. The hope is that GMRES might produce CG like acceleration without squaring the condition number. This is actually true when \( A \) is SPD however, in this case, one should obviously apply CG since it involves much less computation. Even in the case when \( A \) satisfies (15.1), at least theoretically, GMRES may not lead to any computational advantage. Indeed, consider applying CG to the normal equations,

\[
(15.5) \quad A^*Ax = A^*b.
\]

Here \( A^* \) denotes the adjoint of \( A \) with respect to \( < \cdot, \cdot > \). By the definition of \( A^* \), \( A^*A \) is self adjoint with respect to \( < \cdot, \cdot > \). To estimate the convergence associated with conjugate gradient applied to (15.5), we need to estimate the condition number of \( A^*A \). We first observe that
\[
\|Ax\|^2 = < Ax, Ax > \leq \beta \|x\| \|Ax\|
\]
from which it follows that
\[
< A^*Ax, x > = \|Ax\|^2 \leq \beta^2 \|x\|^2.
\]
Thus, the largest eigenvalue \( \lambda_n \) of \( A^*A \) satisfies
\[
\lambda_0 = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{< A^*Ax, x >}{< x, x >} \leq \beta^2.
\]
For the smallest eigenvalue \( \lambda_0 \), we start with the following fact:

\[
(15.6) \quad \|z\| = \sup_{y \in \mathbb{R}^n, y \neq 0} \frac{< z, y >}{\|y\|}.
\]
This inequality is an easy consequence of the Schwarz inequality. Taking \( z = Ax \) gives
\[
\|Ax\| = \sup_{y \in \mathbb{R}^n, y \neq 0} \frac{\langle Ax, y \rangle}{\|y\|} \\
\geq \frac{\langle Ax, x \rangle}{\|x\|} \geq \alpha \|x\|.
\]
Thus,
\[
\lambda_1 = \inf_{x \in \mathbb{R}^n, x \neq 0} \frac{\langle Ax, Ax \rangle}{\|x\|^2} \geq \alpha^2.
\]
It follows that the condition number of \( A \) is bounded by \( \beta^2/\alpha^2 \). Thus, the error \( e_i \) resulting from CG applied to the normal equations satisfies
\[
\|Ae_i\| \leq 2 \left( 1 - \frac{\alpha/\beta}{1 + \alpha/\beta} \right)^i \|Ae_0\|.
\]
We see that CG provides an acceleration over GMRES (at least in theory) converging in \( O(\beta/\alpha) \) iterations as opposed to the \( O(\beta^2/\alpha^2) \) suggested by the above theorem.

We finish this class with a proof of the last two lemmas of the previous class.

**Lemma 1.** Assume that the Arnoldi algorithm does not stop before the \( l \)'th step. Then the vectors \( v_1, v_2, \ldots v_l \) form a \( \langle \cdot, \cdot \rangle \)-orthonormal basis for the Krylov space \( K_l(A) \).

**Proof.** By the definition of \( v_1 \) and (5), \( v_i \) is unit sized for \( i = 1, \ldots, l + 1 \). Provided that \( v_1, \ldots, v_j \) are orthonormal, for \( l = 1, 2, \ldots, j, \)
\[
\langle v_{j+1}, v_l \rangle = h_{j+1,j}^{-1} \langle Av_j, v_l \rangle - \sum_{i=1}^{j} h_{i,j} v_i, v_l > = h_{j+1,j}^{-1} (\langle Av_j, v_l > - h_{i,j}) = 0.
\]
This is the key computation in the induction showing that \( v_1, \ldots, v_{i+1} \) are orthogonal. Hence \( v_1, \ldots, v_{i+1} \) are orthonormal and span an \( l+1 \) dimensional space. A simple induction shows that \( v_i \in K_i \). As \( K_i \) has at most dimension \( i \), \( v_1, \ldots, v_i \) is an orthonormal basis for \( K_i \), for \( i = 1, 2, \ldots, l + 1 \). \( \square \)

**Lemma 2.** Assume that the Arnoldi algorithm does not stop before the \( l \)'th step. For \( i = 1, \ldots, l, \)
\[
Av_i = \sum_{k=1}^{i+1} (H_l)_{k,i} v_k.
\]
Proof. From (2) of the Arnoldi algorithm,

\[ w_j = h_{j+1,j}v_j = Av_j - \sum_{i=1}^{j} h_{i,j}v_i. \]

The lemma follows by rearrangement and the definition of \( \tilde{H} \). \( \square \)

**Example 1.** We consider a finite element or finite difference method for the boundary value problem:

\[-\Delta u + v(x) \cdot \nabla u + q(x)u = f \quad \text{for} \quad x \in \Omega, \]
\[ u = 0 \quad \text{for} \quad x \in \partial \Omega. \]

Here \( v \) is a vector field (a velocity field). The first term above involves higher order derivatives (order 2) while the remaining terms involve lower order derivatives (order 1 and 0, respectively). As we have already seen, discretization of the first term, e.g., with finite differences, leads to a symmetric and positive definite matrix \( A \) while the discretization of the full equation gives a non-symmetric matrix (and possibly no longer definite) \( \tilde{A} \). Often, it is possible to develop an efficient preconditioner for \( \tilde{A} \) satisfying the inequalities:

\[ \alpha \|x\|^2_A \leq (\tilde{B}Ax, x)_A, \quad \text{for all} \quad x \in \mathbb{R}^n, \]
\[ (B Ax, y)_A \leq \beta \|x\|_A \|y\|_A, \quad \text{for all} \quad x, y \in \mathbb{R}^n, \]

with constants \( \alpha, \beta \) independent of \( h \) and hence \( n \). The techniques for deriving such preconditioners and why they lead to the above inequalities is beyond the scope of this class. Note that Theorem 1 can be applied to the preconditioned equations and guarantees a uniform (independent of \( h \)) rate of iterative convergence. This result holds for the preconditioned GMRES method defined using \( \langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_A \). This convergence rate is not guaranteed for the preconditioned GMRES algorithm based on the standard (dot) inner product.