We start this class by investigating the use of $CG$ to get bounds on the spectrum for $A$. As before, $A$ is a SPD $n \times n$ real matrix. We start by considering the Rayleigh-Ritz eigenvalue approximation. For a symmetric matrix $A$,

\begin{equation}
\lambda_n = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{(Ax, x)}{(x, x)} \quad \text{and} \quad \lambda_1 = \min_{x \in \mathbb{R}^n, x \neq 0} \frac{(Ax, x)}{(x, x)}.
\end{equation}

Given a subspace $K \subseteq \mathbb{R}^n$, we can approximate $\lambda_n$ and $\lambda_1$ by replacing $\mathbb{R}^n$ above by $K$, i.e.,

\begin{equation}
\tilde{\lambda}_n = \max_{x \in K, x \neq 0} \frac{(Ax, x)}{(x, x)} \quad \text{and} \quad \tilde{\lambda}_1 = \min_{x \in K, x \neq 0} \frac{(Ax, x)}{(x, x)}.
\end{equation}

This is the so-called Rayleigh-Ritz approximation.

We can get eigenvalue approximations from our CG algorithm by computing $\tilde{\lambda}_n$ and $\tilde{\lambda}_1$ satisfying (13.2) using $K = K_m$ (the $m$'th Krylov space). The numbers $\tilde{\lambda}_n$ and $\tilde{\lambda}_1$ are, respectively, the largest and smallest eigenvalues of the generalized eigenvalue problem: Find pairs $\{\lambda, \phi\}$ with $\phi \in K_m$ and $\lambda \in \mathbb{R}$ satisfying

\begin{equation}
(A\phi, \theta) = \lambda(\phi, \theta), \quad \text{for all } \theta \in K_m.
\end{equation}

It is always possible to solve eigenvalue problems of the form of (13.3) by introducing a basis $\{v_1, \ldots, v_l\}$. Here $l$ denotes the dimension of $K_m$ which may be less then $m$. The obvious basis is $r_0, Ar_0, \ldots, A^{l-1}r_0$ but this turns out not to be the most computationally effective. Given a basis $\{v_1, v_2, \ldots, v_{l-1}\}$, we then solve the (generalized) matrix eigenvalue problem

\begin{equation}
Mx = \lambdaNx \quad \text{for} \quad x \in \mathbb{R}^l.
\end{equation}

Here $M_{i,j} = (Av_i, v_j)$ and $N_{i,j} = (v_i, v_j), \quad i, j = 1, \ldots, l$.

Note that both $M$ and $N$ are SPD $l \times l$ matrices. Although there are routines available for solving this problem, we can use some of the properties of the CG algorithm to develop a much more efficient approach.

Specifically, we use $\{r_0, r_1, \ldots, r_{l-1}\}$ for our basis. Note that by (I.1) in the proof of Theorem (CG-equivalence) $\{p_0, p_1, \ldots, p_{l-1}\}$ is a basis for $K_l$ and so it follows from $p_i = r_i - \beta_{i-1}r_{i-1}$ that $\{r_0, r_1, \ldots, r_{l-1}\}$ is also a
basis for $K_i$. We first note that if $j > i$, (I.2) (again, the proof of Theorem (CG-equivalence) implies

$$(r_j, r_i) = (e_j, r_i)_A = 0$$

so the $N$ matrix is diagonal. Its diagonal entries are given by

$$N_{i,i} = (r_i, r_i) = (r_i, p_i) + \beta_{i-1}(e_i, p_{i-1})_A = (r_i, p_i).$$

Note that the quantity $(r_i, p_i)$ appears as the numerator in $\alpha_i$.

We now consider the matrix $M$. First we observe that it is tridiagonal since if $j > i + 1$, (I.2) implies

$$(Ar_j, r_i) = (e_j, Ar_i)_A = 0.$$ 

We next compute its diagonal entries. Clearly, $M_{1,1} = (Ap_0, p_0)$. This is the denominator in $\alpha_0$. In addition, $r_i = p_i + \beta_{i-1}p_{i-1}$ is an $A$-orthogonal decomposition. The Pythagorean Theorem holds (check it!) for arbitrary inner products so

$$\|r_i\|_A^2 = \|p_i\|_A^2 + \beta_{i-1}^2\|p_{i-1}\|_A^2.$$ 

This can be rewritten

$$M_{i,i} = (Ar_i, r_i) = (Ap_i, p_i) + \beta_{i-1}^2(Ap_{i-1}, p_{i-1}).$$

Thus, for $i > 1$, the value of $M_{i,i}$ can computed from $\beta_{i-1}$ and the denominators for $\alpha_i$ and $\alpha_{i-1}$. Finally, applying (I.1) gives

$$M_{j,j-1} = (Ar_j, r_{j-1}) = (p_j, r_{j-1})_A + \beta_{j-1}(p_{j-1}, r_{j-1})_A$$

$$= \beta_{j-1}(p_{j-1}, r_{j-1})_A = \beta_{j-1}(Ap_{j-1}, p_{j-1}).$$

We see the denominator of $\alpha_{i-1}$ appearing here as well.

Finally, if $x$ satisfies (13.4) with eigenvalue $\lambda$ if and only if $y = N^{1/2}x$ satisfies

$$\widetilde{M}y = \lambda y$$

where $\widetilde{M} = N^{-1/2}MN^{-1/2}$.

This is a standard eigenvalue problem involving a symmetric tridiagonal matrix. The nonzero element of $M$ and $N$ can be gathered during the CG iteration and the matrix $\widetilde{M}$ and its eigenvalues can be computed as part of a post processing step. Efficient procedures for computing the eigenvalues for symmetric tridiagonal matrices are available in standard software libraries such as LAPACK, see www.netlib.org/lapack/.

This technique goes over to the preconditioned case as well and produces estimates for the condition number for the preconditioned system. These estimates are useful in that they provide information on the quality of the preconditioner. Specifically, a small condition number indicates a well designed preconditioner.
Another interesting property of the conjugate gradient method is that it performs well when there are a few outlying eigenvalues with most of the spectrum restricted to a small interval. We illustrate this by a simple example where there is only one outlying eigenvalue. We consider a SPD matrix $A$ with eigenvalues $0 < \lambda_0 << \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Here we assume that $K_1 = \lambda_n / \lambda_1$ is much smaller than $K_0 = \lambda_n / \lambda_0$. Applying the convergence theorem for CG gives

$$\|e_i\|_A \leq \left(\frac{\sqrt{K_0} - 1}{\sqrt{K_0} + 1}\right)^i \|e_0\|_A. \tag{13.5}$$

This estimate assumes nothing about the distribution of the spectrum of $A$.

Alternatively, as conjugate gradient is the minimizer, we can bound the error by approximating it by any element in the Krylov space. In this case, we use

$$\|e_i\|_A \leq \left\| (I - \lambda_0^{-1}A) \prod_{j=1}^{i-1}(I - \tau_jA)e_0 \right\|_A.$$

Here the parameters $\tau_1, \ldots, \tau_{i-1}$ are chosen using the Chebyshev technique applied to the interval $[\lambda_1, \lambda_n]$. We estimate the operator norm on the right hand side above by

$$\left\| (I - \lambda_0^{-1}A) \prod_{j=1}^{i-1}(I - \tau_jA) \right\|_A$$

$$\leq \max_{k=1}^n \left\{ (1 - \lambda_0^{-1}\lambda_k) \prod_{j=1}^{i-1}(I - \tau_j\lambda_k) \right\}$$

$$\leq \max_{\lambda \in [\lambda_1, \lambda_n]} \left\{ (1 - \lambda_0^{-1}\lambda) \prod_{j=1}^{i-1}(I - \tau_j\lambda) \right\}$$

Note that the factor $I - \lambda_0^{-1}\lambda$ produces 0 when $\lambda = \lambda_0$ and hence we can exclude $\lambda_0$ from the middle maximum above. Applying the multi-parameter theorem and the bound

$$|1 - \lambda_0^{-1}\lambda| \leq K_0 \quad \text{for} \quad \lambda \in [\lambda_1, \lambda_n]$$

gives

$$\|e_i\|_A \leq 2K_0 \left(\frac{\sqrt{K_1} - 1}{\sqrt{K_1} + 1}\right)^i \|e_0\|_A. \tag{13.6}$$

Comparing (13.5) and (13.6), we see that although (13.6) has a larger constant ($2K_0$), it produces a much better asymptotic (as $i$ becomes) reduction assuming that $\lambda_1 / \lambda_0$ is large.
MINRES: There are many other Krylov based methods around. For example, we consider the case when $A$ is nonsingular and symmetric (but not positive definite). In this case, $A^2$ is SPD and we consider the minimization problem:

$$x_i = x_0 + \theta \text{ for } \theta \in K_i \equiv K_i(A, r_0)$$

satisfying

$$\|e_i\|_{A^2} = \min_{\zeta \in K_i} \|x - (x_0 + \zeta)\|_{A^2}.$$  

Defining the residual, $r(\zeta) = b - A(x_0 + \zeta)$, (13.7) can be rewritten

$$\|r_i\|_{\ell^2} = \|r(\theta)\|_{\ell^2} = \min_{\zeta \in K_i} \|r(\zeta)\|_{\ell^2}$$

i.e., $x_i$ is the vector in $x_0 + K_i$ which minimizes the residual over all vectors in $x_0 + K_i$.

This problem could be solved by setting up an $i \times i$ matrix as discussed in the lemma of the previous class but can be much more efficiently solved by the following conjugate gradient like algorithm.

**Algorithm 1.** (MINRES). Let $A$ be a symmetric nonsingular $n \times n$ matrix and $x_0 \in \mathbb{R}^n$ (the initial iterate) and $b \in \mathbb{R}^n$ (the right hand side) be given. Start by setting $p_0 = r_0 = b - Ax_0$. Then for $i = 0, 1, \ldots$, define

$$x_{i+1} = x_i + \alpha_i p_i,$$

where

$$\alpha_i = \frac{(r_i, Ap_i)}{(Ap_i, Ap_i)}$$

$$r_{i+1} = r_i - \alpha_i Ap_i,$$

$$p_{i+1} = r_{i+1} - \beta_i p_i,$$

where

$$\beta_i = \frac{(Ar_{i+1}, Ap_i)}{(Ap_i, Ap_i)}.$$  

**Exercise 1.** The above algorithm requires two evaluations of $A$ per step, namely $Ap_i$ and $Ar_{i+1}$. By possibly introducing extra vectors, find an implementation which only requires one $A$ evaluation per step after startup.

That the MINRES algorithm solves the minimization problem (13.8) can be seen by repeating the proof of Theorem (CG-equivalence) replacing the $A$ inner product by the $A^2$-inner product (go through the proof and see!).

**Application:** We consider a finite difference approximation to a Helmholtz equation:

$$-\Delta u(x) - k^2 u(x) = f(x), \text{ in } \Omega,$$

$$u(x) = 0 \text{ on } \partial \Omega.$$  

This equation models, for example, time harmonic acoustic waves in bounded waveguide ($\Omega$ contained in $\mathbb{R}^2$ or $\mathbb{R}^3$). The finite difference equations have two contributions, one from the Laplacian and the other from the lower order term (the $k^2$-term), i.e. $A = A_1 + A_2$. We considered the finite difference approximation to the Laplacian earlier in this course and saw that
it resulted in a symmetric and positive definite matrix $A_1$. In contrast, the second term results in $A_2 = -k^2I$ where $I$ is the identity matrix. Let $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $A_1$. We then have

$$\sigma(A) = \{\lambda_i - k^2 : i = 1, \ldots, n\}.$$  

When $k^2$ is greater than $\lambda_0$, $A$ is symmetric but not positive definite. For this problem, $A$ is nonsingular only if $k^2$ is not an eigenvalue of $A_1$. The resulting matrix $A$ is a natural candidate for the MINRES algorithm.

We next provide an analysis of the above algorithm. As MINRES produces the solution of the minimization (13.7), its analysis consists of coming up with a good approximation in the Krylov space. Recall that for the analysis of CG, we obtained a good approximation in the Krylov space by using the approximation based on the Chebyshev polynomials. We shall use the Chebyshev polynomials again but on an interval which results from the SPD operator $A^2$. Our assumptions on $A$ imply that there are eigenvalues of $A$, namely, $-a, -b, c, d$ with $a, b, c, d$ positive and satisfying

$$\sigma(A) \subset [-a - b] \cap [c, d].$$

The spectrum of $A^2$ is contained in the interval $[\lambda_1, \lambda_n]$ where $\lambda_1 = \min(c^2, b^2)$ and $\lambda_n = \max(a^2, d^2)$. After $2m + 1$ iterations,

$$\|e_{2m+1}\|_{A^2} \leq \left\| \prod_{i=1}^{m} (I - \tau_i A^2) \right\|_{A^2} \|e_0\|_{A^2}$$

where $\{\tau_1, \tau_2, \ldots, \tau_m\}$ are the optimal Chebyshev iteration parameters based on the interval $[\lambda_1, \lambda_n]$. Note that $\lambda_1$ and $\lambda_n$ are, respectively, the smallest and largest eigenvalue of $A^2$ (why?). Applying the multi-parameter theorem gives the following result.

**Theorem 1.** Suppose that $A$ is a symmetric and nonsingular real matrix. Then, the errors corresponding to the MINRES iteration satisfy

$$\|e_{2m+1}\|_{A^2} \leq 2 \left( \frac{\sqrt{K} - 1}{\sqrt{K} + 1} \right)^m \|e_0\|_{A^2}.$$  

Here $K$ is the spectral condition number of $A^2$.

**Remark 1.** Alternatively, one could apply the conjugate gradient method directly to the system $A^2x = Ab$. These are the normal equations in this case. The resulting iteration requires 2 matrix vector evaluations per step (so $m$ steps involve more or less the same number of matrix evaluations as $2m$ steps of MINRES). The resulting cg errors satisfy

$$\|e_m\|_{A^2} \leq 2 \left( \frac{\sqrt{K} - 1}{\sqrt{K} + 1} \right)^m \|e_0\|_{A^2}$$
which is essentially the same as (13.10). Thus, at least theoretically, MINRES is no better than the application of conjugate gradient to the normal equations. Note that we only used the terms in the Krylov space having even powers of $A$ in the analysis of MINRES. The odd powers might give improved convergence in some examples. However, it is possible to construct problems (how would you do this?) for which MINRES and the normal equation approach gives rise to exactly the same sequence of iterates, i.e., the result at $2i + 1$ for MINRES equals the result at $i$ for CG-normal.