STAT 310: MATHEMATICAL COMPUTATIONS II
WINTER 2012
PROBLEM SET 3

In the following, we will let $\langle \cdot, \cdot \rangle$ be an inner product on $\mathbb{R}^n$ and $\|\cdot\|$ be the norm induced. If $A \succ 0$, we write $\langle \mathbf{x}, \mathbf{y} \rangle_A = \mathbf{x}^\top A \mathbf{y}$ and $\|\mathbf{x}\|_A^2 = \mathbf{x}^\top A \mathbf{x}$. The special case $A = I$ will be denoted $\langle \mathbf{x}, \mathbf{y} \rangle_2$ and $\|\mathbf{x}\|_2$ as usual. Recall that for a subspace $V \le \mathbb{R}^n$, its orthogonal complement is $V^\perp := \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{v} \rangle_2 = 0 \text{ for all } \mathbf{v} \in V\}$.

1. Let $\{0\} = V_0 \subset V_1 \subset V_2 \subset \cdots \subset \mathbb{R}^n$ be an increasing sequence of subspaces of $\mathbb{R}^n$. In pure math parlance, such a sequence is often called a *flag* or a *filtration*. Suppose we would like to solve $A\mathbf{x} = \mathbf{b}$ where $A \succ 0$. Let $\mathbf{x}^*$ be its unique solution. Consider the following methods

Orthogonal residual: For $k \in \mathbb{N}$, let $\mathbf{w}_k \in V_k$ be such that

$$A\mathbf{w}_k - \mathbf{b} \in V_k^\perp. \tag{1.1}$$

Minimal residual: For $k \in \mathbb{N}$, let $\mathbf{x}_k \in V_k$ be such that

$$\mathbf{x}_k \in \operatorname*{argmin}_{\mathbf{x} \in V_k} \|A\mathbf{x} - \mathbf{b}\|_2. \tag{1.2}$$

Minimal A-error: For $k \in \mathbb{N}$, let $\mathbf{y}_k \in V_k$ be such that

$$\mathbf{y}_k \in \operatorname*{argmin}_{\mathbf{y} \in V_k} \|\mathbf{y} - \mathbf{x}^*\|_A. \tag{1.3}$$

Minimal A^{-1}-residual: For $k \in \mathbb{N}$, let $\mathbf{z}_k \in V_k$ be such that

$$\mathbf{z}_k \in \operatorname*{argmin}_{\mathbf{z} \in V_k} \|A\mathbf{z} - \mathbf{b}\|_{A^{-1}}. \tag{1.4}$$

(a) Which of $\mathbf{w}_k$, $\mathbf{x}_k$, $\mathbf{y}_k$ and $\mathbf{z}_k$ as defined above are unique?
(b) Which of $\mathbf{w}_k$, $\mathbf{x}_k$, $\mathbf{y}_k$, $\mathbf{z}_k$ as defined above are equal for all $k \in \mathbb{N}$?
(c) Show that if A is only nonsingular, then (a) and (b) are false in general but $\mathbf{w}_k$ and $\mathbf{x}_k$ will still converge to $\mathbf{x}^*$.

2. Let $A \in \mathbb{R}^{n \times n}$. Recall that a subspace $V \le \mathbb{R}^n$ is said to be A-invariant if

$$A(V) \subseteq V,$$

i.e. $A\mathbf{v} \in V$ for all $\mathbf{v} \in V$. Let $K_k(A, \mathbf{b}) = \operatorname{span}\{\mathbf{b}, A\mathbf{b}, \dots, A^{k-1}\mathbf{b}\}$ be the kth Krylov subspace.
(a) Let A be nonsingular. Prove that the following statements are equivalent.
  (i) $\mathbf{b}, A\mathbf{b}, \dots, A^k\mathbf{b}$ are linearly dependent.
  (ii) $K_k(A, \mathbf{b}) = K_{k+1}(A, \mathbf{b})$.
  (iii) $K_k(A, \mathbf{b})$ is A-invariant.
  (iv) There exists an A-invariant subspace $V \le \mathbb{R}^n$ with $\dim(V) \le k$ and $\mathbf{b} \in V$.
  (v) $A^{-1}\mathbf{b} \in K_k(A, \mathbf{b})$.
(b) Show that $K_n(A, \mathbf{b})$ is the intersection of all A-invariant subspaces in $\mathbb{R}^n$ containing $\mathbf{b}$.
(c) Show that

$$K_\infty(A, \mathbf{b}) = K_n(A, \mathbf{b}) = K_m(A, \mathbf{b})$$

where m is the degree of the minimal polynomial of A.

Date: January 29, 2012 (Version 1.4); due: February 3, 2012.

(d) Recall that the Arnoldi algorithm quits at step j if $h_{j+1,j} = 0$. Show that if this happens and A is nonsingular, then
$$A^{-1}\mathbf{b} \in K_j(A, \mathbf{b}).$$

3. In general, if we apply Gram-Schmidt algorithm to linearly independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ to obtain orthonormal $\mathbf{q}_1, \dots, \mathbf{q}_n$, we get a k-term recurrence
$$r_{kk}\mathbf{q}_k = \mathbf{a}_k - \sum_{i=1}^{k-1} r_{ik}\mathbf{q}_i \tag{3.5}$$
where
$$r_{ik} = \mathbf{q}_i^\top \mathbf{a}_k \qquad \text{and} \qquad |r_{kk}| = \left\| \mathbf{a}_k - \sum_{i=1}^{k-1} r_{ik}\mathbf{q}_i \right\|_2.$$

(a) Show that when applied to the Krylov sequence $\mathbf{b}, A\mathbf{b}, \dots, A^{k-1}\mathbf{b}$ where $A \succ 0$, we always get a three-term recurrence in (3.5). How is this related to Lanczos algorithm?

(b) Consider the continuous analogue of this process where we apply Gram-Schmidt to $f_1, \dots, f_n \in L^2[-1, 1]$. Here the matrix $A \in \mathbb{R}^{n\times n}$ is replaced by the linear operator $A : L^2[-1, 1] \to L^2[-1, 1]$ defined by
$$(Af)(x) = xf(x)$$
for any $f \in L^2[-1, 1]$, i.e. A is the 'multiplication by x' operator. Write down a continuous analogue of the Lanczos algorithm for this choice of A.

(c) Show that when applied to the Krylov sequence with $f(x) = 1$ playing the role of $\mathbf{b}$, we get
$$xq_k(x) = \beta_{k-1}q_{k-1}(x) + \alpha_k q_k(x) + \beta_{k+1}q_{k+1}(x)$$
where $\beta_0 = 0$ and $q_0(x) = 0$ and for $k \in \mathbb{N}$,
$$\alpha_k = 0, \qquad \beta_k = \frac{1}{2}\left(1 - \frac{1}{4k^2}\right)^{-1/2}.$$
These are known as the Legendre polynomials[1] and could also be obtained by applying Gram-Schmidt to the Krylov sequence $1, x, x^2, x^3, \dots$ directly.

4. Suppose we would like to solve $A\mathbf{x} = \mathbf{b}$ where $A \succ 0$ using the conjugate gradient method.

(a) Show that the sequence of iterates and residuals generated by the conjugate gradient satisfies
$$\mathbf{x}_k = p_{k-1}(A)\mathbf{b}, \quad \mathbf{r}_k = q_k(A)\mathbf{b}$$
where $p_{k-1} \in \mathbb{R}[x]$ has degree $k - 1$ and
$$q_k(x) = xp_{k-1}(x) - 1.$$

(b) Show that if
$$p_{k-1}(t) = c_0 + c_1 t + \cdots + c_{k-1}t^{k-1},$$
then the coefficients $(c_0, c_1, \dots, c_{k-1})^\top \in \mathbb{R}^k$ is a solution of the following linear system
$$\begin{bmatrix} \mathbf{b}^\top A\mathbf{b} & \mathbf{b}^\top A^2\mathbf{b} & \cdots & \mathbf{b}^\top A^k\mathbf{b} \\ \mathbf{b}^\top A^2\mathbf{b} & \mathbf{b}^\top A^3\mathbf{b} & \cdots & \mathbf{b}^\top A^{k+1}\mathbf{b} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}^T A^k\mathbf{b} & \mathbf{b}^\top A^{k+1}\mathbf{b} & \cdots & \mathbf{b}^\top A^{2k+1}\mathbf{b} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{k-1} \end{bmatrix} = \begin{bmatrix} \mathbf{b}^\top\mathbf{b} \\ \mathbf{b}^\top A\mathbf{b} \\ \vdots \\ \mathbf{b}^\top A^{k-1}\mathbf{b} \end{bmatrix}.$$

---

[1] We could also get other orthogonal polynomials if we choose an inner product like
$$\langle f, g\rangle_w = \int_{-1}^{1} f(x)g(x)w(x)\,dx$$
with an appropriate choice of weight function w. Here we used $w(x) = 1$ to get the Legendre polynomials. Had we used $w(x) = 1/\sqrt{1 - x^2}$, we would have obtained the Chebyshev polynomials that we discussed last time.