Math 270A: Numerical Linear Algebra

Instructor: Michael Holst

Fall Quarter 2014

Homework Assignment #3
Due Give to TA at least a few days before final if you want feedback.

Exercise 3.1. (The Basic Linear Method for Linear Systems)

1. Recall that the transpose A^T of an $n \times n$ matrix A is defined as

$$A^T_{ij} = A_{ji}.$$

Another characterization of the transpose matrix A^T is that it is the unique *adjoint* operator satisfying

$$(Au, v) = (u, A^T v), \quad \forall u, v \in \mathbb{R}^n,$$

where $(\cdot, \cdot)$ is the usual Euclidean inner-product,

$$(u, v) = \sum_{i=1}^{n} u_i v_i.$$

Show that these two characterizations are equivalent.

2. Given an SPD (symmetric positive definite) $n \times n$ matrix A, show that it defines a new inner-product

$$(u, v)_A = (Au, v) = \sum_{i=1}^{n} (Au)_i v_i, \quad \forall u, v \in \mathbb{R}^n,$$

called the A-inner-product. I.e., show that $(u, v)_A$ is a "true" inner-product, in that it satisfies the three inner-product properties.

3. The A-adjoint of a matrix M, denoted M^*, is defined as the adjoint in the A-inner-product; i.e., the unique matrix satisfying:

$$(AMu, v) = (Au, M^* v), \quad \forall u, v \in \mathbb{R}^n.$$

Show that that an equivalent definition of M^* is

$$M^* = A^{-1} M^T A.$$

4. Now that we have a new inner-product $(u, v)_A$, give a definition of the norm it induces, which we will denote $\|u\|_A$. Give an argument that it satisfies the three norm properties.

5. Use the vector norm $\|u\|_A$ to define the operator (matrix) norm $\|M\|_A$ which it induces for any $n \times n$ matrix M. Give an argument that this definition gives us access to the following inequality:

$$\|Mu\|_A \leq \|M\|_A \|u\|_A, \quad \forall u \in \mathbb{R}^n.$$

6. Use the A-norm to define the A-condition number of any square $n \times n$ matrix M, which we denote as $\kappa_A(M)$.

7. Prove the following: If $M \in \mathbb{R}^{n \times n}$ is A-self-adjoint for an SPD $A \in \mathbb{R}^{n \times n}$, then

$$\rho(M) = \|M\|_A, \qquad \kappa_A(M) = \frac{|\lambda_{\max}(M)|}{|\lambda_{\min}(M)|}.$$

8. Let $A, B \in \mathbb{R}^{n \times n}$ be SPD, and let $\alpha \in \mathbb{R}$ with $\alpha \neq 0$. Our goal is to solve the linear system $Au = f$ for the unknown $u \in \mathbb{R}^n$, given $f \in \mathbb{R}^n$, and given $B \approx A^{-1}$. Derive the following identity and argue that it has a unique solution $u \in \mathbb{R}^n$ that agrees with the solution to $Au = f$:

$$u = (I - \alpha BA)u + \alpha Bf.$$

9. Derive the parameterized form of the Basic Linear Method for the problem $Au = f$:

$$u^{n+1} = (I - \alpha BA)u^n + \alpha Bf.$$

10. Use the Banach Fixed Point Theorem to prove that the BLM converges if α is restricted to a subset of $\mathbb{R}$, and give that subset.
11. Prove that when the optimal value is chosen for α, the following error bound holds for the Basic Linear Method:

$$\|e^{n+1}\| \leq \left(1 - \frac{2}{1 + \kappa_A(BA)}\right)^{n+1} \|e^0\|,$$

where $e^{n+1} = u - u^{n+1}$, and where $\kappa_A(BA)$ denotes the A-condition number of BA.
12. Drive a more general error bound for the Basic Linear Method, similar to the result above, but one that holds for any value of α. (Hint: The norm of the error propagator $E = I - BA$ will appear more naturally than the condition number of BA.)
13. Assume we have the optimal value for α. Assuming that we would like to achieve the following accuracy in our iteration after some number of steps n:

$$\frac{\|e^{n+1}\|_A}{\|e^0\|_A} < \epsilon,$$

use the approximation:

$$\ln\left(\frac{a-1}{a+1}\right) = \ln\left(\frac{1 + (-1/a)}{1 - (-1/a)}\right) = 2\left[\left(\frac{-1}{a}\right) + \frac{1}{3}\left(\frac{-1}{a}\right)^3 + \frac{1}{5}\left(\frac{-1}{a}\right)^5 + \cdots\right] < \frac{-2}{a},$$

to show that we can achieve this error tolerance with the Basic Linear Method if n satisfies:

$$n = O\left(\kappa_A(BA)\,|\ln \epsilon|\right).$$

14. Derive a similar iteration bound result that holds for any α.
15. Many types of matrices have $O(1)$ non-zeros per row (the finite difference and finite element discretizations of ordinary and partial differential equations we looked at briefly in 270A, and will look at more in 270C, always generate such matrices). Assume that the cost to store and apply B is $O(n)$. Prove that the cost of one iteration of BLM will be $O(n)$, where n is the dimension of the problem, i.e., A is an $n \times n$ matrix.
16. What is the overall complexity (in terms of n and $\kappa_A(BA)$) to solve the problem with the BLM to a given tolerance ϵ? (Assume you have the optimal α.) If $\kappa_A(BA)$ can be bounded by a constant, independent of the problem size n, what is the complexity? Is this then an optimal method?

Exercise 3.2. (Derivation of the Conjugate Gradient Method.)
In this problem, we will derive the conjugate gradient (CG) method, an extremely important iterative method for solving matrix equations $Au = f$ when A is symmetric ($A = A^T$) and positive-definite ($u^T Au > 0, \quad \forall u \neq 0$). (We will refer to matrices having both properties as SPD.) This problem brings together several different topics we have discussed in lecture in 270A, including iterative methods for matrix equations, manipulations with inner-products and norms in a vector space, together with some basic ideas from approximation and orthogonal polynomials that will be examined more completely in 270B.

1. The Cayley-Hamilton Theorem states that a square $n \times n$ matrix M satisfies its own characteristic equation:
$$P_n(M) = 0.$$
Using this result, prove that if M is also nonsingular, then the matrix M^{-1} can be written as a matrix polynomial of degree $n-1$ in M, or
$$M^{-1} = Q_{n-1}(M).$$

2. Consider now the matrix equation:
$$Au = f,$$
where A is an $n \times n$ SPD matrix, and u and f are n-vectors. It is common to "precondition" such an equation before attempting numerical solution, by multiplying by an approximate inverse operator $B \approx A^{-1}$ and then solving the *preconditioned system*:
$$BAu = Bf.$$
If A and B are both SPD, under what conditions is BA also SPD? Show that if A and B are both SPD, then BA is A-SPD (symmetric and positive in the A-inner-product).

3. Given an initial guess u^0 to the solution of $BAu = Bf$, we can form the initial residuals:
$$r^0 = f - Au^0, \qquad s^0 = Br^0 = Bf - BAu^0.$$
Do a simple manipulation to show that the solution u can be written as:
$$u = u^0 + Q_{n-1}(BA)s^0,$$
where $Q(\cdot)$ is the matrix polynomial representing $(BA)^{-1}$. In other words, you have established that the solution u lies in a *translated Krylov space*:
$$u \in u^0 + V_{n-1}(BA, s^0),$$
where
$$V_{n-1}(BA, s^0) = span\{s^0, BAs^0, (BA)^2 s^0, \ldots, (BA)^{n-1} s^0\}.$$
Note that we can view the Krylov spaces as a sequence of expanding subspaces
$$V_0(BA, s^0) \subset V_1(BA, s^0) \subset \cdots \subset V_{n-1}(BA, s^0).$$

4. We will now try to construct an iterative method (the CG method) for finding u. The algorithm determines the best approximation u_k to u in a subspace $V_k(BA, s^0)$ at each step k of the algorithm, by forming
$$u^{k+1} = u^k + \alpha_k p^k,$$
where p^k is such that $p^k \in V_k(BA, s^0)$ at step k, but $p^k \notin V_j(BA, s^0)$ for $j < k$. In addition, we want to enforce minimization of the error in the A-norm,
$$\|e^{k+1}\|_A = \|u - u^{k+1}\|_A,$$
at step k of the algorithm. The next iteration expands the subspace to $V_{k+1}(BA, s^0)$, finds the best approximation in the expanded space, and so on, until the exact solution in $V_{n-1}(BA, s^0)$ is reached.
To realize this algorithm, let's consider how to construct the required vectors p^k in an efficient way. Let $p^0 = s^0$, and consider the construction of an A-orthogonal basis for $V_{n-1}(BA, s^0)$ using the standard Gram-Schmidt procedure:
$$p^{k+1} = BAp^k - \sum_{i=0}^{k} \frac{(BAp^k, p^i)_A}{(p^i, p^i)_A} p^i, \qquad k = 0, \ldots, n-2.$$

At each step of the procedure, we will have generated an A-orthogonal (orthogonal in the A-inner-product) basis $\{p^0, \ldots, p^k\}$ for $V_k(BA, s^0)$. Now, note that by construction,

$$(p^k, v)_A = 0, \quad \forall v \in V_j(BA, s^0), \quad j < k.$$

Using this fact, and the fact you established previously that BA is A-self-adjoint, show that the Gram-Schmidt procedure has only three non-zero terms in the sum, namely

$$p^{k+1} = BAp^k - \frac{(BAp^k, p^k)_A}{(p^k, p^k)_A} p^k - \frac{(BAp^k, p^{k-1})_A}{(p^{k-1}, p^{k-1})_A} p^{k-1}, \qquad k = 0, \ldots, n-1.$$

Thus, there exists an efficient three-term recursion for generating the A-orthogonal basis for the solution space. Note that this three-term recursion is possible due to the fact that we are working with orthogonal (matrix) polynomials.

5. We can nearly write down the CG method now, by attempting to expand the solution in terms of our cheaply generated A-orthogonal basis. However, we need to determine how far to move in each "conjugate" direction p^k at step k after we generate p^k from the recursion. As remarked earlier, we would like to enforce minimization of the quantity

$$\|e^{k+1}\|_A = \|u - u^{k+1}\|_A$$

at step k of the iterative algorithm. Show that this is equivalent to enforcing

$$(e^{k+1}, p^k)_A = 0.$$

(I.e., show that minimizing the error in finding an approximation from a subspace is mathematically equivalent to enforcing that the approximation error be orthogonal to that subspace.)

6. Let's assume we have somehow enforced

$$(e^k, p^i)_A = 0, \quad i < k,$$

at the previous step of the algorithm. We have at our disposal $p^k \in V_k(BA, s^0)$, and let's take our new approximation at step $k+1$ as:

$$u^{k+1} = u^k + \alpha_k p^k,$$

for some step-length $\alpha_k \in \mathbb{R}$ in the direction p^k. Thus, the error in the new approximation is simply:

$$e^{k+1} = e^k + \alpha_k p^k.$$

Show that in order to enforce $(e^{k+1}, p^k)_A = 0$, we must choose α_k to be

$$\alpha_k = \frac{(r^k, p^k)}{(p^k, p^k)_A}.$$

7. The Conjugate Gradient Algorithm you have derive above is:

(The Conjugate Gradient Algorithm)
Let $u^0 \in \mathcal{H}$ be given.
$r^0 = f - Au^0, \quad s^0 = Br^0, \quad p^0 = s^0.$
Do $k = 0, 1, \ldots$ until convergence:
$\quad \alpha_k = (r^k, p^k)/(p^k, p^k)_A$
$\quad u^{k+1} = u^k + \alpha_k p^k$
$\quad r^{k+1} = r^k - \alpha_k A p^k$
$\quad s^{k+1} = Br^{k+1}$
$\quad \beta_{k+1} = -(BAp^k, p^k)_A/(p^k, p^k)_A$
$\quad \gamma_{k+1} = -(BAp^k, p^{k-1})_A/(p^{k-1}, p^{k-1})_A$
$\quad p^{k+1} = BAp^k + \beta_{k+1} p^k + \gamma_{k+1} p^{k-1}$
End do.

Show that equivalent expressions for some of the parameters in CG are:

(a) $\alpha_k = (r^k, s^k)/(p^k, p^k)_A$
(b) $\delta_{k+1} = (r^{k+1}, s^{k+1})/(r^k, s^k)$
(c) $p^{k+1} = s^{k+1} + \delta_{k+1} p^k$

Remark: The CG algorithm that appears in textbooks is usually formulated to employ these equivalent expressions due to the reduction in computational work of each iteration.

Exercise 3.3. (Properties of the Conjugate Gradient Method.)

In this problem, we will establish some simple properties of the CG method derived in Problem 1. (Although this analysis is standard, you will have difficulty finding all of the pieces in one text.)

1. Show that the error in the CG algorithm propagates as:

$$e^{k+1} = [I - BAp_k(BA)]e^0,$$

where $p_k \in \mathcal{P}_k$, the space of polynomials of degree k.

2. By construction, we know that the polynomial above is such that

$$\|e^{k+1}\|_A = \min_{p_k \in \mathcal{P}_k} \|[I - BAp_k(BA)]e^0\|_A.$$

Now, since BA is A-SPD, we know that it has real positive eigenvalues $\lambda_j \in \sigma(BA)$, and further, that the corresponding eigenvectors v_j of BA are orthonormal. Using the expansion of the initial error

$$e^0 = \Sigma_{j=1}^n \alpha_j v_j,$$

establish the inequality:

$$\|e^{k+1}\|_A \le \left(\min_{p_k \in \mathcal{P}_k} \left[\max_{\lambda_j \in \sigma(BA)} |1 - \lambda_j p_k(\lambda_j)| \right] \right) \|e^0\|_A.$$

The polynomial which minimizes the maximum norm above is said to solve a *mini-max problem.*

3. It is well-known in approximation theory that the Chebyshev polynomials

$$T_k(x) = \cos(k \arccos x)$$

solve mini-max problems of the above type, in the sense that they deviate least from zero (in the *max-norm* sense) in the interval $[-1, 1]$, which can be shown is due to their unique *equi-oscillation* property. (These facts can be found in any introductory numerical analysis text, such as the one used in Math 170C.) If we extend the Chebyshev polynomials outside the interval $[-1, 1]$ in the natural way (outlined in class), it be shown that shifted and scaled forms of the Chebyshev polynomials solve the above mini-max problem. In particular, the solution is simply

$$1 - \lambda p_k(\lambda) = \tilde{p}_{k+1}(\lambda) = \frac{T_{k+1}\left(\frac{\lambda_{max}+\lambda_{min}-2\lambda}{\lambda_{max}-\lambda_{min}}\right)}{T_{k+1}\left(\frac{\lambda_{max}+\lambda_{min}}{\lambda_{max}-\lambda_{min}}\right)}.$$

Use an obvious property of the polynomials $T_{k+1}(x)$ to conclude

$$\|e^{k+1}\|_A \le \left[T_{k+1}\left(\frac{\lambda_{max}+\lambda_{min}}{\lambda_{max}-\lambda_{min}} \right) \right]^{-1} \|e_0\|_A.$$

4. Use one of the Chebyshev polynomial results given in the last problem below to refine this inequality to:

$$\|e^{k+1}\|_A \le 2 \left(\frac{\sqrt{\frac{\lambda_{max}(BA)}{\lambda_{min}(BA)}} - 1}{\sqrt{\frac{\lambda_{max}(BA)}{\lambda_{min}(BA)}} + 1} \right)^{k+1} \|e^0\|_A.$$

5. Using the fact that BA is A-self-adjoint and positive definite, show that the above error reduction inequality can be written more simply as

$$\|e^{k+1}\|_A \le 2 \left(\frac{\sqrt{\kappa_A(BA)} - 1}{\sqrt{\kappa_A(BA)} + 1} \right)^{k+1} \|e^0\|_A = 2 \left(1 - \frac{2}{1 + \sqrt{\kappa_A(BA)}} \right)^{k+1} \|e^0\|_A.$$

6. Assume that we would like to achieve the following accuracy in our iteration after some number of steps n:

$$\frac{\|e^{n+1}\|_A}{\|e^0\|_A} < \epsilon.$$

Using the approximation:

$$\ln\left(\frac{a-1}{a+1}\right) = \ln\left(\frac{1+(-1/a)}{1-(-1/a)}\right) = 2\left[\left(\frac{-1}{a}\right) + \frac{1}{3}\left(\frac{-1}{a}\right)^3 + \frac{1}{5}\left(\frac{-1}{a}\right)^5 + \cdots\right] < \frac{-2}{a},$$

show that we can achieve this error tolerance if n satisfies

$$n = O\left(\kappa_A^{1/2}(BA)\left|\ln\frac{\epsilon}{2}\right|\right).$$

7. Many types of matrices have $O(1)$ non-zeros per row (the finite difference and finite element discretizations of ordinary and partial differential equations we look at next quarter always generate such matrices). Therefore, the cost of one iteration of CG will be $O(n)$, where n is the dimension of the problem, i.e., A is an $n \times n$ matrix. What is the overall complexity (in terms of n and $\kappa_A(BA)$) to solve the problem to a given tolerance ϵ? If $\kappa_A(BA)$ can be bounded by a constant, independent of the problem size n, what is the complexity? Is this then an optimal method?
8. As noted earlier, many types of matrices have $O(1)$ non-zeros per row (the finite difference and finite element discretizations of ordinary and partial differential equations we looked at briefly in 270A, and will look at more in 270C, always generate such matrices). Assume that the cost to store and apply B is $O(n)$. Prove that the cost of one iteration of CG will be $O(n)$, where n is the dimension of the problem, i.e., A is an $n \times n$ matrix.
9. What is the overall complexity (in terms of n and $\kappa_A(BA)$) to solve the problem with CG to a given tolerance ϵ? If $\kappa_A(BA)$ can be bounded by a constant, independent of the problem size n, what is the complexity? Is this then an optimal method?

Exercise 3.4. (Properties of the Chebyshev Polynomials.)

The Chebyshev polynomials are defined as:

$$t_n(x) = \cos(n\cos^{-1} x), \qquad n = 0, 1, 2, \ldots.$$

Taking $t_0(x) = 1$, $t_1(x) = x$, it can be show that the Chebyshev polynomials are an orthogonal family that can be generated by the standard recursion (which holds for any orthogonal polynomial family):

$$t_{n+1}(x) = 2t_1(x)t_n(x) - t_{n-1}(x), \qquad n = 1, 2, 3, \ldots.$$

Prove the following extremely useful relationships:

$$t_k(x) = \frac{1}{2}\left[\left(x + \sqrt{x^2-1}\right)^k + \left(x - \sqrt{x^2-1}\right)^k\right], \qquad \forall x, \tag{3.1}$$

$$t_k\left(\frac{\alpha+1}{\alpha-1}\right) > \frac{1}{2}\left(\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}-1}\right)^k, \qquad \forall \alpha > 1. \tag{3.2}$$

(These two results are fundamental in the convergence analysis of the conjugate gradient method in the earlier parts of the homework.)

Hint: For the first result, use the face that $\cos k\theta = (e^{ik\theta} + e^{-ik\theta})/2$. The second result will follow from the first after a lot of algebra.