M143M

Watkins' reference
+ study guides

FO9  LF
References in Watkins's text (Fundamentals of Matrix Computations)

Prior to the midterm:

Linear Algebra Review and Norms:
Sections: 1.1, 1.2, 2.1

Basic Decomposition:
LU decomposition: sections 1.7
QR decomposition: sections 3.1, 3.2, 3.4
Eigenvalue decomposition: sections: 5.2
Singular Value Decomposition: sections: 4.1 and pp. 275-278

Gaussian Elimination:
Gaussian elimination and its stability: sections 1.3, 1.7, 1.8
Operation counts: pp. 3-7 and each section where an algorithm is discussed
Stability: section 2.7
Inverses and Determinant: pp. 103-104

Cholesky and SPD matrices:
diagonal dominance: see class notes (alternate reference Numerical Linear Algebra and Application by Biswath Datta pp. 21-23, pp. 228-230)
SPD matrices and Cholesky algorithm: section 1.4

Condition Numbers and Computer Arithmetic:
Condition numbers: sections 2.2
Computer arithmetic: section 2.5, 2.6

after the midterm:

Rectangular Linear Systems (Least Squares), Gram-Schmidt and Householder:
Sections 3.1, 3.2, 3.3, 3.4

SVD and Sensitivity of Least Squares Problems
Sections 4.1, 4.2, 4.3, 4.4

Eigenvalues:
Introduction and power method – sections 5.1, 5.2, 5.3
Similarity transformations – section 5.4
QR algorithm for eigenvalues – section 5.6

Iterative Methods:
Preconditioner – section 7.5
Conjugate gradient method – section 7.6
Krylov subspace method = section 7.9

Software:
Appendix A
General
1. Name 4 conditions that are equivalent to the existence of a unique solution to \( Ax = b \).
2. Solve a lower or upper triangular system.
3. Count the operations of an algorithm like these.
4. When is a triangular matrix singular?

SPD Matrices
5. What are some conditions that imply that matrix is symmetric positive definite?
6. Be able to prove simple related results. (For example prove that if A is spd then A is nonsingular or that if \( A = MM^T \) with M nonsingular then A is spd).
7. What is diagonal dominance and how does it relate to spd matrices?
8. Carry out the Cholesky decomposition via the border methods.
9. Write out these algorithms for a general spd matrix A.
10. Use a picture relating to matrix multiplication to motivate the algorithm.
11. How is Cholesky’s decomposition used to solve \( A x = b \)?

Gaussian elimination algorithms
12. What are the algorithms for genp (no pivoting) and gepp (partial pivoting) for a general matrix A?
13. Use these to solve specific examples
14. Draw a picture relating to step k of Gaussian elimination and use it to describe the difference between partial pivoting and complete pivoting.
15. Be able to describe some alternative pivoting schemes (for example rook pivoting) to these above “standard” schemes.

Operation counts
16. Know the operation counts below and how to derive the counts for the first four algorithm listed.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>flops (+, *, -, / or ( \sqrt{} ))</th>
<th>comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>triangular solve</td>
<td>( n^2 )</td>
<td>0</td>
</tr>
<tr>
<td>Cholesky (for spd)</td>
<td>( n^3/3 )</td>
<td>0</td>
</tr>
<tr>
<td>genp</td>
<td>( 2 \frac{n^3}{3} )</td>
<td>0</td>
</tr>
<tr>
<td>gepp</td>
<td>( 2 \frac{n^3}{3} )</td>
<td>( (1/2) n^2 )</td>
</tr>
<tr>
<td>gepp</td>
<td>( 2 \frac{n^3}{3} )</td>
<td>( (1/3) n^3 )</td>
</tr>
<tr>
<td>ge with rook pivoting</td>
<td>( 2 \frac{n^3}{3} )</td>
<td>( (2.7 / 2) n^2 ), typically</td>
</tr>
</tbody>
</table>

17. Know why based on the above counts that gepp is preferred over gepp.

Accuracy
18. What is swamping and how is it related to subtractive cancellation in Gaussian elimination with no pivoting (genp)? Is genp numerically stable?
19. If \( \hat{x} \) is the calculated solution to \( Ax = b \) using Gaussian elimination and \( \hat{x} \) satisfies \( (A + \delta A) \hat{x} = b \), what is a formula that bounds \( \| \delta A \| / \| A \| \)?
20. What is the "growth factor" for Gaussian elimination and why is it important?
21. What is a theoretical bound on the growth factor for gepp and gecp? How fast does each bound grow with n?
22. In practice, for gepp, is the theoretical bound on the growth factor achieved? Discuss.
23. Know why gepp is used in practice even though the theoretical bound for error growth is better for gecp than gepp.
24. What is the growth factor for the Cholesky decomposition of an spd matrix. If A is spd, is the Cholesky decomposition stable?

Miscellaneous
25. Why is using the LU decomposition better than using "old fashion" Gaussian elimination (i.e. tacking b onto A and operating on the bigger matrix).
26. When is A = LU possible? Give an example where it doesn't work.
27. Be able to prove related results (For example prove that if A = L U, with L unit lower triangular and U nonsingular upper triangular then det(A_k) is not zero).
28. When is PA = LU possible with L unit lower triangular and U nonsingular upper triangular?
29. For large n, describe a good algorithm and a poor algorithm for finding det(A).
30. For large n, what is an algorithm for finding A^{-1}?
31. For large n, what is the disadvantage of solving Ax = b via x = A^{-1} b?
32. What happens to gepp if it is applied to a singular matrix?
33. Know why a block algorithm is important and how cache memory is significant on modern computers.
34. What are LAPACK and LINPACK and what is their key difference.
35. What are the BLAS and why are they important. What effect does using machine language (rather than Fortran) BLAS have on an algorithm.

MCS 143M – F09                         Test 1 Study Guide: Miscellaneous

General
1. Know the basic matrix decompositions and one example of what they are used for: A = QR (QR factorization) for least squares, A = U D V^T (singular value decomposition) for matrix approximation, PA = LU (LU factorization) for solving square linear systems for example in approximate solutions to a differential equation, and A = VDV^{-1} (eigenvalue) for identifying resonance frequencies. Know why resonance is important in practice.

Approximate solution to differential equations
2. Know how to start with a differential equation and approximate it with a linear system of equations.
4. Know that the approximation and therefore the calculated solution gets better as the step size gets smaller and the number of equations gets larger.
Norms
1. What are three properties that characterize a vector norm?
2. Define some commonly used norms such as the 1, 2, p, and infinity norm.
3. Be able to prove that these are norms.
4. What are four properties that characterize a matrix norm?
5. What is an induced matrix norm?
6. How does one calculate a Frobenius, one and infinity norm of a matrix?
7. What is a formula for a matrix 2-norm and why is this difficult to calculate?
8. Be able to prove simple properties of these matrix norms.
9. Define the singular value decomposition. How does it relate to matrix norms?
10. What is the two norm of an orthogonal matrix?

Condition Numbers
11. In solving $Ax=b$ how are the relative errors in $x$ related to the relative errors in $b$?
12. How is this relation derived?
13. If the $\text{cond}(A)$ and the number of significant digits in $b$ are known be able to determine a bound on the number of significant digits in a calculated solution $x$ to $Ax = b$.
14. Be able to calculate the condition number with the one or infinity norm.
15. Be able to prove results concerning condition numbers (e.g.: $\text{cond}(cA) = \text{cond}(A))$.
16. What is the classic example of an ill conditioned matrix?
17. For a 2 by 2 system draw a picture corresponding to an ill conditioned system. Describe why small changes in the system can cause large changes in the solution.
18. Is a small determinant a good indication that $A$ is nearly singular? Provide an example to justify your answer.
19. How is $\text{cond}(A)$ related to how close $A$ is to being singular?

Computer Arithmetic
20. What is an overflow error? What is an underflow error? What are typical values of each in IEEE single and double precision?
21. What is roundoff error?
22. What is relative machine precision and how is it related to roundoff error?
23. Why is the IEEE standard for floating point arithmetic important? What are some ways the IEEE standard handles special cases?
24. Know our definition of a the numerical stability of an an algorithm: an algorithm is numerically stable if a bound on the relative error in the algorithm is close to (a small or modest multiple) of the bound on the relative error inherent in the problem.
25. What is the difference between a numerically unstable algorithm and an ill condition problem? Give an example of each.
26. Describe ways that might be used to cure a numerically unstable algorithm.
27. Describe ways that might be used to cure an ill conditioned problem.
28. What is subtractive cancellation and why is it important?
29. Provide an example of a simple formula that has subtractive cancellation that can be avoided by rewriting the formula.
30. Be able to identify subtractive cancellation in an expression and rewrite it to avoid subtractive cancellation. Know how to do this using algebra, trig identities and Taylor's series.
Iterative Methods Study Guide

General
1. What is a sparse matrix and why is the concept important?
2. Describe the non-zero structure of the matrix that arises when using the five point star discretization with Laplace's equation over a rectangle.
3. Be able to construct the matrix $A$ and right hand side $b$ for a simple problems involving the five point star discretization with Laplace's equation over a rectangle.
4. Know that for linear systems arising from 3 or higher dimensional partial differential equations, sparse Gaussian elimination may not compete, in terms of efficiency, with iterative methods.

CG
5. Know how and why the solution to $Au = b$ is related to $\min F(u) = 0.5 u^T A u - b^T u$ for symmetric positive definite matrices $A$
6. Why, for spd matrices $A$, does $F(u)$ decrease (or not increase) at each step of conjugate gradients?
7. What property of the residuals can be used to motivate the term "conjugate" in the cg method?
8. What is a three-term recurrence and why is it important for the efficiency of the method?
9. Know that cg requires $10n$ flops plus one multiplication of the form $Ax$ per step.
10. In principle what is the maximum number of step required by the cg method? Why?
11. Be able to carry out a step or two of the cg method for a 2 by 2 matrix.
12. If $A$ has condition number $\kappa$, what is a formula for the bound on the error after $m$ steps of using the cg method? For what matrices $A$ (what values of $\kappa$) will the cg method converge quickly? Slowly? Why?
13. What is the preconditioned conjugate gradient method? How does it relate to the ordinary cg method? Why should it converge faster?
14. What two goals do you want to satisfy in choosing a preconditioner?

GMRES
15. Know that gmres satisfies a minimization problem that guarantees (in exact arithmetic) that the calculated solution improves (or gets no worse) from step to step.
16. Does gmres have a three term recurrence relation?
17. Know that the work per step increases as the number of iterations increases.
18. How does the restarted gmres method address the issue in 17.

BICG
20. Know that bicg does not solve a minimization problem at each step and that there is no guarantee that the calculated solution improves. The convergence can fail and / or be erratic.
21. Know that bicg does satisfy a three-term recurrence and that the work per step does not increase as the number of iterations increases.

QMR, BICGSTAB and CGS
22. Know that each of these tries to improve bicg. Cgs tries to avoid use of $A^T$. Bicgstab and QMR try to improve the reliability and stability of bicg.
23. These methods do not solve a minimization problem and that there is no guarantee that the calculated solution improves. The convergence can fail and/or be erratic. However QMR and biesttab use modifications of biccg that, in some cases, avoids erratic convergence.

24. Know that these methods satisfy three term recurrences and that the work per step does not increase as the number of iterations increases.

**For all methods**

25. How does the cond(A) affect the convergence rate of the method?

26. Know that preconditioning can speed up the convergence, sometimes dramatically.

27. Know that in exact arithmetic, the methods should converge in a number of steps $\leq$ the dimension of A, but to be of practical value they must converge much sooner than this.

**Miscellaneous**

28. Know that there are older methods – Jacobi, Gauss-Sidel and successive over relaxation (sor) that we did not cover. Jacobi and Gauss-Sidel are typically slower than the methods we discussed. Sor can be competitive with the methods discussed in some cases.
Application
1. Be able to setup a least squares problem given data points in the plane and a model with unknown coefficients that is supposed to fit the data points.
2. Know what a Vandermonde matrix is.

Orthogonal matrices
3. What is an orthogonal matrix?
4. Be able to prove elementary properties of an orthogonal matrix such as: the product of orthogonal matrices is orthogonal, the inverse of an orthogonal matrix is its transpose, and multiplication by an orthogonal matrix does not change 2 norms.
5. Prove that the solution to a least squares problem with matrix $A$ and right hand side $b$ is unchanged when an orthogonal matrix is applied to $A$ and $b$.

Normal equations
6. Draw a picture that describes the range of a rectangular matrix $A$.
7. Derive the normal equations, making use of a picture that indicates how being perpendicular (or normal) is important in a least squares problem.
8. Use normal equation to find the solution to a least squares problem.
9. Describe two ways that the solution to a least squares problem with a rectangular matrix $A$ ($m$ by $n$ with $m \geq n$) can be converted to the solution to a square system: by normal equations and by orthogonal transformations.
10. Know that normal equations can, in some cases, introduce more computer arithmetic errors than introduced by the better orthogonal matrix methods (SVD or Householder).

GS
11. Know that the basic idea of Gram-Schmidt: for each new vector one finds its projection onto the space perpendicular to all the preceding vectors.
12. Know that GS can introduce more computer arithmetic errors than does Householder’s method or a variation, which we did not discuss, Modified Gram-Schmidt (MGS).

svd
13. Be able to solve a least squares problem given the singular value decomposition (svd) of $A$.
14. Know how to use the singular value decomposition to approximate a matrix. Know that the approximation can be very good when there is a large “gap” in the singular values of $A$.
15. Know how to use the svd to find the (absolute value of the) determinant of a matrix, the inverse of a matrix and the two norm of a matrix.
16. Know the relationship of the svd to the eigenvalues and eigenvectors of $A^T A$ and $A A^T$.
17. Know that the singular values of a diagonal matrix are the absolute values of the diagonal entries.
18. Be able to find the singular value decompositions of matrices that are modifications of diagonal matrices.
19. Know that the condition number of square or rectangular matrix ($m$ by $n$ with $m \geq n$) is the ratio of the first singular value of $A$ over the $n$th singular value of $A$. Such a matrix is singular or has “infinite condition number” if its $n$th singular value is zero.

Householder
20. Prove that an elementary Householder transform $U = 1 - 2 u u^T$ is orthogonal if $u^T u = 1$.
21. Given vectors $v$ and $w$, construct a matrix $U$ that corresponds to the reflection that moves $v$ into $(-w)$. (Hint the normal to the “mirror” will be in direction $\|v\| w + \|w\| v$.)
22. Be able to draw (at least in $\mathbb{R}^2$) the “mirror” that corresponds to the reflection in the preceding point.

23. Describe how a sequence of reflection can be used to triangularize a matrix.

24. Be able to solve a least squares problem with a 3 by 2 matrix by Householder’s method.

25. Know that Householder’s idea is based on reflections. Know that one can also base an algorithm on rotations, called Given’s rotations. We did not discuss Given’s rotations since Given’s rotations are less efficient than on Householder reflections.

**Efficiency**

26. When deriving operation counts of algorithms know how to approximate summations with integrals.

27. Be able to derive operation counts for algorithms involving matrices: for example multiplication of two general matrices, multiplication by triangular matrices, or multiplication of a matrix by an elementary Householder transform.

28. Know the following counts for number of operations ($x$, $+$, $-$, $/$, or square root) when solving least squares problems:

<table>
<thead>
<tr>
<th>Method</th>
<th>$m \geq n$</th>
<th>$m \equiv n$</th>
<th>$m \gg n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>$mn^2 + n^3 / 3$</td>
<td>$(4/3)n^3$</td>
<td>$mn^2$</td>
</tr>
<tr>
<td>GS or MGS</td>
<td>$2mn^2$</td>
<td>$2n^3$</td>
<td>$2mn^2$</td>
</tr>
<tr>
<td>Householder</td>
<td>$2mn^2 - (2/3)n^3$</td>
<td>$(4/3)n^3$</td>
<td>$2mn^2$</td>
</tr>
<tr>
<td>svd</td>
<td>$2mn^2 + (2/3)n^3$</td>
<td>$(8/3)n^3$</td>
<td>$2mn^2$</td>
</tr>
<tr>
<td>Givens</td>
<td>$4mn^2 - (4/3)n^3$</td>
<td>$(8/3)n^3$</td>
<td>$4mn^2$</td>
</tr>
</tbody>
</table>

29. Which method(s) are more efficient for $m \equiv n$?

30. Which method is more efficient for $m \gg n$?

**Accuracy**

31. Define the condition number of a rectangular matrix. (see 19, above)

32. Know that when you use normal equations the relative error growth is always proportional to the square of $\text{cond}(A)$.

33. Know that if one uses orthogonal based methods (svd or Householder) then the relative error growth depends on the size of the residual $r = b - Ax$. If $r$ is small then the error growth is proportional to $\text{cond}(A)$. If $r$ is large then the error growth is proportional to the square of $\text{cond}(A)$.

34. Know a formula for the bound on the relative error when using an algorithm based on normal equations.

35. If the condition of $A$ is small (roughly one) how does the accuracy of normal equations and Householder compare?

36. If the condition of $A$ is moderately large, how does the accuracy of normal equations and orthogonal methods compare if the residual is small? If the residual is moderately small? If the residual is large?
Resonance
1. What is resonance?
2. For an oscillating string or other structure what is a physical interpretation of eigenvalues?
3. For an oscillating string or other structure what is a physical interpretation of eigenvectors?
4. Given a partial differential equation \( u_t(s,t) - k(x) u_{xx}(x,t) = g(x,t) \) for \( a <= x <= b \) be able to write a matrix whose eigenvectors and eigenvalues have physical meaning.
5. What effect does increasing the size of the matrix have on the calculated eigenvalues and eigenvectors?
6. Which matrix eigenvalues lead to the better approximations to the true resonance frequencies?

Definitions and standard theorems
7. Know that the eigenvalue decomposition is \( A = V D V^{-1} \) with \( V \) nonsingular and \( D \) diagonal
8. What are similar matrices and how are the eigenvalues of the matrices related?
9. If \( A \) is symmetric know that eigenvalues are real and eigenvectors are orthogonal. \( A \)
10. If \( A \) is nonsymmetric but real know that eigenvalues are real or come in complex conjugate pairs.
11. Know that a normal matrix has \( A^H A = A A^H \). Here the superscript \( H \) indicates complex conjugate transpose.
12. Know that \( A \) symmetric \( \Rightarrow \) \( A \) is normal and \( A \) is normal \( \Rightarrow \) that \( A = V D V^{-1} \) with \( V \) unitary and \( D \) diagonal.
13. Know that eigenvalues can be calculated accurately for normal matrices.
14. Know that if \( \lambda \) is an eigenvalue of \( A \), \( \hat{\lambda} \) is the closest eigenvalue of \( \hat{A} \) to \( \lambda \) and \( V \) is the matrix of eigenvectors then \( |\lambda - \hat{\lambda}| \leq \text{cond}(V) \|A - \hat{A}\| \). Therefore if \( A \) is far from normal (in the sense that \( \text{cond}(V) \) is far from 1) then it can be difficult to calculate eigenvalues accurately.
15. Know that in the case for matrices far from normal it is better to use the Schur decomposition \( A = Q R Q^{-1} \), where \( Q \) is unitary and \( R \) triangular rather than the eigenvalue decomposition.
16. Know that for some matrices \( A \) it is not possible to write \( A \) as \( A = V D V^{-1} \) with \( V \) nonsingular and \( D \) diagonal. \( A \) is not diagonalizable and is called defective.

Calculation of eigenvalues
17. For a general 5 by 5 or larger matrix it is not possible to have an algorithm that will find the eigenvalues exactly in a finite number of steps. Why not? Iterative methods required to calculate matrices that are 5 by 5 or larger?
18. Be able to apply the power method to calculate an eigenvalue and eigenvector.
19. What is the convergence rate of the power method?
20. Know that basic idea of the QR algorithm for eigenvalues is to construct a sequence of matrices similar to \( A \) using \( A_i = QR \) and \( A_{i+1} = RQ \) and that the sequence converges to a triangular matrix.
21. What are the eigenvalues of a triangular matrix?
22. Know that calculation of all the eigenvalues of a matrix requires, in practice, \( O(n^3) \) flops.