Matrix Theory
Review for Final Exam

The final exam is Wednesday, December 13 at 8:00-10:00 in our usual classroom. The exam will cover the topics listed below.

Eigenvalues and Eigenvectors
Let A be an n by n matrix. Then an eigenvector of A is a nonzero vector x such that Ax and x are multiples of each other. In otherwords $Ax = \lambda x$ for some scalar λ. An eigenvalue of A is a scalar λ so that $Ax = \lambda x$ for some nonzero vector x. An eigenpair of A is a pair (λ, x) where λ is a scalar, and x is a nonzero vector, such that $Ax = \lambda x$. Be able to prove things about eigenvectors and eigenvalues. Geometrically, an eigenvector is a direction that A stretches, and an eigenvalue is a stretching factor of A. It is possible to show that eigenvalues of A corresponding to distinct eigenvectors are linearly independent. Thus, in particular, an n by n matrix has at most n distinct eigenvalues.

Orthogonal sets, projections and least squares
A set of nonzero vectors is orthogonal if $u \cdot v = 0$ for each pair of vectors in the set, and is orthonormal if in addition the vectors each have length 1. Any set of orthogonal (nonzero) vectors is linearly independent. This can be argued as follows: Suppose that $u_1, \ldots, u_k$ is an orthogonal set of vectors. Suppose that $\vec{0} = c_1u_1 + c_2u_2 + \cdots + c_ku_k$. Dotting with u_i gives: $0 = c_iu_i \cdot u_i$. Since $u_i \neq \vec{0}$, $u_i \cdot u_i \neq 0$. So each $c_i = 0$, and the vectors are linearly independent.

A subspace of $\mathbb{C}^n$ (or $\mathbb{R}^n$) is a set U of vectors that is closed under addition (which means if $u, v \in U$, then $u + v \in U$) and closed under scalar multiplcation (which means whenever $u \in U$ and k is a scalar, the vector ku belongs to U). Examples of subspaces of $\mathbb{R}^4$ are the origin, lines through the origin, planes through the origin, hyperplanes through the origin, and all of $\mathbb{R}^4$.

A common problem is: given a subspace U and a point v determine the point p in U that is closest to v. It turns out that p is the unique point on U so that $v - p$ is orthogonal to each vector in U (i.e. $(v - p) \cdot u = 0$ for each $u \in U$.) To see this, suppose that $v - p$ is orthogonal to each vector in U, and let u be a vector in U. Then the points u, v and p form the vertices of a triangle. The two (non-hypotenuse) sides are $v - p$ and $p - u$. Thus, by the Pythagorean Theorem, $\|v - u\|^2 = \|v - p\|^2 + \|p - u\|^2$. We conclude that the distance between v and u is at least the distance between v and p. Hence, p is the closest point in U to v. The point p is called the projection of v onto U, and is denoted by $\text{proj}_U(v)$.

In the case that U has an orthonormal basis $\alpha = \{w_1, w_2, \ldots, w_k\}$, the projection of v onto U is given by $\text{proj}_U v = (w_1 \cdot v)w_1 + (w_2 \cdot v)w_2 + \cdots (w_k \cdot v)w_k$. This can be verified as follows: We show that $v - [(w_1 \cdot v)w_1 + (w_2 \cdot v)w_2 + \cdots (w_k \cdot v)w_k]$ orthogonal to each vector of U. To do this, simply dot this vector with w_j, and use that $w_1, \ldots, w_k$ are orthonormal. We get $w_j \cdot v - (w_j \cdot v)(w_j \cdot w_j) = w_j \cdot v - w_j \cdot v = 0$. f $v \in U$, then $\text{proj}_U v = v$, and hence $[v]_\alpha = \begin{bmatrix} w_1 \cdot v \\ w_2 \cdot v \\ \vdots \\ w_n \cdot v \end{bmatrix}$.

We see that having an orthogonal basis of a subspace U is important. So we need to be able to find one. We do this via the Gram-Schmidt process. We start with a basis $u_1, u_2, \ldots, u_k$ of U. We define $v_1 = u_1$, and $w_1 = v_1/(\|v_1\|)$. Next we set $v_2 = u_2 - \text{proj}_{u_1} u_2$, and $w_2 = v_2/(\|v_2\|)$. At the jth step, we set $v_j = u_j = \text{proj}_{u_1, \ldots, u_{j-1}} u_j$, and $w_j = v_j/\|v_j\|$. One can show that v_i and v_j are orthogonal when $i \neq j$, as follows: We have that $v_j = u_j - \sum_{k=1}^{n} (u_j \cdot v_k)/(v_k \cdot v_k)$. Dotting with v_i and using that v_i is orthogonal to v_k with $k \neq i$ and $k < j - 1$, we get that $v_j \cdot v_i = u_j \cdot v_i - (u_j \cdot v_i)/(v_i \cdot v_i)(v_i \cdot v_i) = 0$.

In real-life applications, we often run experiments to see how certain data are correlated. This often leads to solving systems of equations of the form $Ax = b$ where A is an m by n matrix, and b is an m by 1 vector. Due to data-collection errors, the system $Ax = b$ invariably has no solutions. This is not acceptable. So what we do, is try to find a point $\hat{b}$ closest to b for which $Ax = \hat{b}$ has a solution, and then use that solution as the probable solution to the original problem. The set $U = \{Ax : x \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$ because $Ax + Ay = A(x + y)$ and $A(kx) = kAx$. Note that $Ax = \hat{b}$ has a solution if and only if $\hat{b} \in U$. Thus, we need to find the projection of b onto U. To do this we seek an x so that $b - Ax$ is orthogonal to each vector in U. Thus, we want an x so that the dot-product with each column of A and $(b - Ax)$ is 0. This leads to $A^*(b - Ax) = \vec{0}$, because (the entries of $A^{ast}(b - Ax)$ record the dot-products between columns of A and $b - Ax$. In otherwords, to find the "best" solution to $Ax = b$, we solve the normal equations $A^*Ax = A^*b$. It turns out that in general, the normal equations have a solution. In the case that the columns of A are linearly independent this comes from the fact that A^*A is invertible (this can be argued as follows: $A^*Ax = \vec{0} \to x^*A^*Ax = 0 \to \|Ax\| = 0 \to Ax = \vec{0} \to x = \vec{0}$.). Thus, in this case, the "best" solution to $Ax = b$ is $x = (A^TA)^{-1}A^Tb$. We call x the least squares solution to $Ax = b$ because is minimize the sum of the squares of the elements of $Ax - b$. . Thus, even though it sounds fancy least squares (or regression) is nothing more than finding the point on a subspace closest to a given point.

Be able to explain how to find the least squares solution using normal equations, that is, be able to do a problem like:

> This problem concerns explaining the meaning of the least squares solution to $Ax = b$ where A is an m by n matrix, b is an m by 1 vector, and x is an n by 1 vector of unknowns.
> Let y be a vector so that $A^TAy = A^Tb$. Thus, y is the least squares solution to $Ax = b$.
>
> 1. Show that each vector of the form Az, $z \in R^n$ is orthogonal to $b - Ay$.
> 2. Let U be the subspace of $\mathbb{R}^m$ consisting of the vectors of the form Az, $z \in \mathbb{R}^n$. Use your answer in (a), and the Pythagorean theorem, to show that Ay is the point in U closest to b. (Hint: Relate the distances of the sides of the triangle whose vertices are b, Ay and Az.)
> 3. Explain, using (b), why when working with data from experiments it is reasonable to use y as a solution to $Ax = b$—even when $Ax = b$ has no solutions.

Unitary & Orthogonal matrices

A matrix A is unitary if $A^*A = I$. Each of the following is equivalent to the square matrix being unitary: (a) $A^{-1} = A$, (b) the map that sends x to Ax preserves lengths, (c) the map that sends x to Ax preserves dot-products, (d) $AA^* = I$, (e) the rows of A are an orthonormal basis, and (f) the columns of A are an orthonormal basis.

A real unitary matrix is called an *orthogonal matrix*. The a real matrix A is an orthogonal matrix if and only if each of its columns has length 1 and its columns are pairwise orthogonal, or more simply $A^TA = I$.

Unitary similarity The matrices A and B are unitarily similar if there exists a unitary matrix U such that $U^*AU = B$. In other words, A and B are unitarily similar if and only if A and B are the same but with respect to different orthonormal bases.

Note that if (λ, x) is an eigenpair of B, then (λ, Ux), is an eigenpair of A. Thus, not only are the eigenvalues of unitarily similar matrices the same, but also the angles between eigenvectors (here's why: the cosine of the angle between u and v is $u \cdot v)/\|u\|\|v\|$). But, since U is unitary $Uu \cdots Uv = u \cdot v$, $\|Uu\| = \|u\|$, and $\|Uv\| = \|v\|$. So the angle between u and v is the same as that between Uu and Uv).

A matrix A is symmetric if A is real and $A^T = A$. Eigenvalues and eigenvectors of symmetric matrices have nice properties. In particular, if λ is an eigenvalue of A, then λ is real (because if $Ax = \lambda x$, then $x^*Ax = x^*\lambda x = \lambda x^*x$. Also $x^*Ax = (A^*x)^*x = (Ax)^*x = (\lambda x)^*x = \bar{\lambda}x^*x$. So $\lambda = \bar{\lambda}$, that is, λ is real); and eigenvectors corresponding to distinct eigenvalues of A are orthogonal (here's why: Suppose that (λ, x) and (μ, y) are eigenvectors of the symmetric matrix A. Then $y^*Ax = y^*(Ax) = y^{ast}(\lambda x) = \lambda(y^*x$. Also $y^*Ax == (Ay)^*x = (\mu y)^*x = \mu y^*x$. So $\mu y^*x = \lambda y^*x$. Since $\lambda \neq \mu$, we conclude that $y^*x = 0$, that is, y and x are orthogonal. Using these facts one can show that if A is symmetric, then there exists an orthogonal matrix Q and a diagonal matrix D so that $Q^TAQ = D$. That is, A is orthogonally similar to a diagonal matrix. In essence, this says that the way A acts on R^n is to stretch it in n orthogonal directions.

One can find Q as follows: (a) find the eigenvalues of A; (b) for each eigenvalue of A find an orthonormal set of eigenvectors (you might have to use Gram-Schmidt here); (c) Let Q be the matrix with columns the vectors found in (b).

Normal Matrices

It is nice to know which matrices are unitarily diagonalizable, i.e. for which A does there exist a unitary matrix U and a diagonal matrix D such that $U^*AU = D$. In other words, which A are unitarily, camouflaged diagonal matrices.

Once nice property about diagonal matrices is that they commute with their conjugate transpose, that is, if D is diagonal then $DD^* = D^*D$. Thus we would expect unitarily diagonalizable matrices to commute with their conjugate transpose. Indeed this is true: if $U^*AU = D$, then $AA^* = A^*A$, because $DD^* = D^*D$.

A square matrix A is normal if $AA^* = A^*A$. Examples of normal matrices are: symmetric matrices, diagonal matrices, hermitian matrices, skew-symmetric matrices, skew-hermitian matrices, $\cdots$. The previous paragraph shows that if A is unitarily diagonalizable, then it is normal. The converse to this holds, namely, if A is normal, then A is unitarily diagonalizable. T

Positive semidefinite matrices
A square matrix A is positive semidefinite if A is hermitian and each of its eigenvalues is a nonnegative real number. .

Every matrix A of the form $A = B^*B$ is positive semidefinite because $x^*B^*Bx == \|Bx\|^2 \geq 0$.

A diagonal matrix A is positive semidefinite if and only if each of its diagonal entries is nonnegative.

If A is positive semidefinite, and (λ, x) is an eigenpair, then $\lambda x^*x = x^*Ax \geq 0$. Thus, each eigenvalue of a positive semidefinite matrix is nonnegative.

Suppose A is an hermitian matrix and each of its eigenvalues is nonnegative. Then A is normal (because $A^*A = AA = AA^*$) and there is a unitary U and a diagonal matrix such that $U^*AU = D$. The diagonal entries of D are nonnegative because $D^* = (U^*AU)^* = U^*A^*U = U^*AU = D$. . Let $\sqrt{D}$ be the matrix obtained from D by taking the square root of each diagonal entry. Then $\sqrt{D}\sqrt{D}^* = D = U^*AU$. Solving for A gives $U\sqrt{D}\sqrt{D}^*U^* = A$. Thus, $A = BB^*$, where $B = U\sqrt{D}$.

Just as each nonnegative number has a square root, each positive semidefinite matrix has a (positive semidefinite) square root. Let A be a positive semidefinite matrix. Then A is simply a camouflaged, nonnegative diagonal matrix. That is, $A = UDU^*$ for some unitary matrix A, and some nonnegative diagonal matrix D. Now, D has a positive semidefinite square root (namely, $\sqrt{D}$ the matrix obtained by taking square roots of each of D's diagonals). So naturally, A has a positive semidefinite square roots–namely $B = U\sqrt{D}U^*$. So B is positive semidefinite because (it is hermitian and its eigenvalues (those of $\sqrt{D}$ are nonnegative) , and $B^2 = A$, because $B^2 = U\sqrt{D}U^*U\sqrt{D}U^* = U\sqrt{D}\sqrt{D}U^* = UDU^* = A$.

The Polar Decomposition
Every complex number z can be written as $z = re^{i\theta}$ where r is a nonnegative real number, and $e^{i\theta}$ is a complex number of modulus 1.

Analogously, every n by n matrix A can be written as $A = RU$, where R is positive semidefinite, and U is unitary. To see this (in the case A is invertible), set $R = \sqrt{AA^*}$. Then R is positive semidefinite by definition, and is invertible since A is. Now set $U = R^{-1}A$. Then U is unitary because $U^*U = (R^{-1}A)^*R^{-1}A = A^*(R^{-1})^*R^{-1}A = A^*R^{-2}A = A^*(AA^*)^{-1}A = A^*(A^*)^{-1}A^{-1}A = I \cdot I = I$. Also, $A = RU$ because $RU = R(R^{-1}U) = U$.

The SVD

Theorem 1 *Let A be an m by n matrix. Then there exist unitary matrices U and V and a nonnegative real numbers, $\sigma_1 \geq \sigma_2 \geq \sigma_m \geq 0$. so that $A = U\Sigma V$, where Σ is the m by n matrix with (j,j)-entry equal to σ_j for $j = 1, 2, \ldots, \min(m,n)$, and all other entries equal to* 0.

Note that $A = U\Sigma V$ says that

$$A = \sum_{i=j}^{\min(m,n)} \sigma u_j v_j^*$$

where $u_1, \ldots, u_m$ are the columns of U and $v_1, v_2, \ldots, v_n$ are the columns of V. The factorization $A = U\Sigma V^*$ called the *singular value decomposition*, or *SVD*, of A. Thus, the singular value decomposition says that A can be written as a sum of very simple matrices $\sigma_j u_j v_j^*$, and that $Av_j = \sigma_j u_j$ for $j = 1, 2, \ldots, \min(m,n)$, and $Av_j = \vec{0}$ otherwise.

Note that if $A = U\Sigma V^*$, then $AA^* = U\Sigma^2 U^*$, and it follows that $\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2$ are the eigenvalues of AA^*.

More on the SVD–for your personal edification Given the singular value decomposition $A = \sum_{j=1}^m \sigma_j u_j v_j^*$ and an integer $k \leq m$, we define A_k by $A_k = \sum_{j=1}^k \sigma_j u_j v_j^*$. That is, A_k is obtained from A by just summing the first k terms of the SVD. For an m by n matrix B, we let $\|B\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |b_{ij}|^2}$ denote the "length" of the matrix B. Note $\|B\|^2$ is the sum of the squared lengths of the columns of B, or equivalently the squared lengths of the rows of B. Hence, if U is unitary, then $\|UB\| = \|B\| = \|BU\|$. Also, $\|B - C\|$ denotes the distance between the matrix B and the matrix C. Thus, if we let Σ_k be the m by n matrix obtained from Σ by zeroing out its last $m - k$ diagonal entries, then

$$\|A - A_k| = \|U\Sigma V^* - U\Sigma_k V^*)| = |U(\Sigma - \Sigma_k)V^*| = |\Sigma - \Sigma_k| = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_m^2}.$$

Thus, if $\sqrt{\sigma_{k+1}^2 + \cdots \sigma_m^2}$ is small, then A_k is a good approximation of A.

This is used in image processing. For simplicitly assume that we wish to store in a computer a black and white picture (say m pixels by n pixels). We can store the picture in an m by n matrix $A = [a_{ij}]$ whose (i, j)-entry is 1 if the (i, j)-pixel is white, and is 0 if the (i, j)-pixel is 1. To store A requires mn bits of information. For m and n large, this is too much information to store. So we approximate A by $A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^*$. We don't store A_k, but rather we store the σ_j, u_j and v_j for $j = 1, 2, \ldots, k$. Thus, to store this information that we can reconstruct A_k from costs $k(m+n+1)$ bits. If k is significantly less than m and n (as is frequently the case), then it costs much less to store "A_k" than is does A. However, A_k might be much different than A. In particular, A_k won't have just 0 and 1 entries. However, if $\|A - A_k\| < 1$, then $A = \lceil A_k \rceil$, where $\lceil A_k \rceil$ is the matrix obtained from A_k by taking the greatest integer of each entry of A_k. Thus, we can completely reconstruct A (exactly!) from the σ_j, u_j, and v_j with $j = 1, 2, \ldots, k$. This method is frequently used to compress videos or pictures.

1. Let A be a matrix such that $A = B^*B$ where B is an n by n matrix.

   (a) Show that A is hermitian.
   To show A is hermitian we need to show that $A = A^*$. Now: $A^* = (B^*B)^* = B^*(B^*)^* = B^*B$. So A is hermitian.

   (b) Show that A is positive semidefinite.
   $x^*Ax = x^*B^*Bx = (Bx)^*Bx = \|Bx\|^2 \geq 0$. Since A is hermitian, and x^*Ax is always nonnegative, A is positive semidefinite.

2. Let A be an n by n skew-symmetric matrix.

   (a) Show that A is normal.
   We need to show that $AA^* = A^*A$. Now, A is real and $A^T = -A$, so $A^* = -A$. So $AA^* = -A^2 = A^*A$, and A is normal.

   (b) Let (λ, x) be an eigenpair of A. Show that λ is purely imaginary.
   We know $Ax = \lambda x$. So $x^*Ax = \lambda x^*x$. But $x^*A = (A^*x)^* = (-Ax)^* = (-\lambda x)^* = -\overline{\lambda}x^*$. So $x^*Ax = -\lambda x^*x$. We conclude that $\lambda x^*x = -\overline{\lambda}x^*x$. Since $x \neq \vec{0}$, , $x^*x \neq 0$, and we have $\lambda = -\overline{\lambda}$, and thus, λ is purely imaginary.

   (c) Let (μ, y) be another eigenpair of A with $\mu \neq \lambda$. Show that x and y are orthogonal.
   Consider y^*Ax. On the one hand this is: $y^*(Ax) = y^*(\lambda x) = \lambda y^*x$. On the other hand we have: $y^*Ax = (A^*y)^*x = (-Ay)^*x = (-\mu y)^*x = -\bar{\mu}y^*x = \mu y^*x$. So $\lambda y^*x = \mu y^*x$. Since $\lambda \neq \mu$, we musth that $y^*x = 0$. So x and y are orthogonal.

3. The matrix A is an n by n matrix and Q is an n by n orthogonal matrix whose last column is an eigenvector of A corresponding to the eigenvalue 7.

   (a) Carefully, explain why the last column of Q^TAQ has the form

   $$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 7 \end{bmatrix}$$

   Let x be the last column of Q. Then the last column of Q^TAQ is $Q^TAx = Q^T(7x) = 7Q^Tx$. The ith entry of Q^Tx is the dot-product of the i column of Q and x, and is 0 if $i \neq n$, and is 1 if $i = n$.

   (b) It is known that R is an orthogonal matrix so that R^TBR is lower triangular. Use Q and R to find an orthogonal matrix S so that S^TAS is lower triangular.
   Take $S = (R \oplus [1])Q$. Then S is orthogonal, and S^TAS is lower triangular.

4. It is known that A is a 3 by 3 real matrix such that

$$Q^TAQ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where

$$Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}.$$

It can be verified that Q is an orthogonal matrix.

(a) What are the eigenvalues of A?
Since A is similar to $Q^T AQ$, the eigenvalues of A are those of $Q^T AQ$, which are the diagonal entries, 1, 16 and 0.

(b) For one of the eigenvalues in (a) give a corresponding eigenvalue of Q.
For $\lambda = 1$, an eigenvector of Q is the first column of Q. For $\lambda = 16$, an eigenvector of Q is the second column of Q. For $\lambda = 0$, an eigenvector of Q is the third column of Q.

(c) Is A positive semidefinite?
Yes. $A = Q^T \mathrm{diag}(1, 16, 0)Q$. So, $A^* = A$, and each eigenvalue of A is a nonnegative real number.

(d) Find a positive semidefinite matrix B such that $B^4 = A$.
Take $B = Q \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^T$.

5. Let A be a 5 by 6 matrix of zeros and ones. Its singular value decomposition is

$$A = 2u_1v_1^* + 3u_2v_2^* + 2u_3v_3^* + .4u_4v_4^* + .2u_5v_5^*.$$

(a) Determine Av_2. Since $v_1, v_2, \ldots, v_5$ is an orthonormal set: $Av_2 = 3u_2v_2 \cdot v_2 = 3u_2$.

(b) Determine with justification A^*u_2. $A^{ast} = 2v_1u_1^* + 3v_2u_2^* + 2v_3u_3^* + .4v_4u_4^* + .2v_5u_5^*$. Since $u_1, \ldots, u_5$ is an orthonormal set: $A^*u + 2 = 3v_2$.

(c) Would A_3 be a good approximation to A?
$\|A - A_3\|^2 = .4^2 + .2^2 = .20$.
So A_3 is $\sqrt{.2} < 1/2$ away from A. . Thus, A_3 is a good approximation of A.

(d) Briefly explain how you could use A_3 to recover A.
Since $\|A - A_3\| < 1/2$, each entry of A and A_3 differ by at most 1/2. Since A has only 0's and 1's, the (i, j)-entry of A will simply be the (i, j)-entry of A_3 rounded to the nearest integer.

6. Let A be a 3 by 3 real symmetric matrix, Q an orthonormal matrix such that $Q^T AQ$ is the diagonal matrix

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and b a 3 by 1 vector. We wish to find a vector x so that $\|Ax - b\|$ is minimal.

(a) Explain why for each real vector x we have $\|Q^T Ax - Q^T b\| = \|Ax - b\|$.
Since Q is an orthogonal matrix, so is Q^T. Since pre-multiplication by an orthogonal matrix preserves length, we have $\|Q^T Ax - Q^T b\| = \|Ax - b\|$

(b) Let $y = Q^T x$. Show that $Dy - Q^T b = Q^T x - Q^T b$.
Simply substitute y into the lefthand side to get: $Dy - Q^T b = DQ^T x - Q^T b$

(c) Explain why (a) and (b) give that

$$\min_x \|Ax - b\| = \min_{y:} \|Dy - Q^T b\|$$

Since, Q^T is orthogonal the correspondence $x \iff Q^T x$ sends R^3 to all of R^3. By (b), $\|Ax - b\| = \|Dy - Q^T b\|$.

(d) It is known that

$$Q^T b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Using (c), determine $\min_x \|Ax - b\|$

By (c), we need to minimize $\|Dy - Q^T b\|$. The possibly Dy's are of the form

$$\begin{bmatrix} 2y_1 \\ y_2 \\ 0 \end{bmatrix}.$$

So the closest to $Q^T b$ we can get to is with $y_1 = 1/2$, and $y_2 = 1$. This gives us a distance of 1.

Solutions to Optional Homework

1. (a) $v_1 = u_1$

$$v_2 = u_2 - (\tfrac{u_2 \cdot v_1}{v_1 \cdot v_1})v_1 = u_2 - (2/2)v_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

$$v_3 = u_3 - (\tfrac{u_3 \cdot v_1}{v_1 \cdot v_1})v_1 - (\tfrac{u_3 \cdot v_2}{v_2 \cdot v_2})v_2 = u_3 - (4/2)v_1 - (2/4)v_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

So $w_1 = v_1/\sqrt{2}$, $w_2 = v_2/2$ and $w_3 = v_3/\sqrt{3}$.

(b) $\text{proj}_U v = (v \cdot w_1)w_1 + (v \cdot w_2)w_2 + (v \cdot w_3)w_3 = 0w_1 + 1w_2 + \sqrt{3}w_3 = v$ (Note: this says is that v is in U)

(c) The distance is $|v - \text{proj}_U v\| = 0$.

2. (a) $Ax = 5x$

(b) y isn't a multiple of x, so x and $y$are linearly independent. z can't be built out of x and y, so x, y, z are linearly independent vectors of $\mathbb{R}^3$. Every set of linearly independent vectors in $\mathbb{R}^3$ is a basis of $\mathbb{R}^3$.

(c) We'll get $R = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{bmatrix}$

(d) The $(j,1)$-entry of $R^T AR$ is the jth entry of $R^T Ax$, which equals that jth entry of $5R^T x$ (since $Ax = 5x$). The jth entry of $5R^T x$ is 5 times the dot-product of the jth column of R and x. The dot-product is 1 if $j = 1$ and is 0 otherwise. So the $(j,1)$ entrys of $R^T AR$ is 5 is $j = 1$ and is 0 otherwise.

(e) This follows from (e)

(f) So $R^T AR$ has the form

$$\left[\begin{array}{c|cc} 5 & u & v \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{B}} \\ 0 & & \end{array}\right]$$

for some 2 by 2 matrix B (Note: $R^T AR$ need not be symmetric. If A is symmetric, then $R^T AR$ is symmetric.)

(g) $B \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. We can take

$$S = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

(h)

$$\hat{S}^T R^T AR\hat{S} = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{S^T} \\ 0 & & \end{array}\right] \left[\begin{array}{c|cc} 5 & u & v \\ \hline 0 & \multicolumn{2}{c}{B} \\ 0 & & \end{array}\right] \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{S^T} \\ 0 & & \end{array}\right]$$

$$= \left[\begin{array}{c|c} 5 & \begin{bmatrix} u & v \end{bmatrix} S \\ \hline 0 & \multirow{2}{*}{S^T BS} \\ 0 & \end{array}\right] = \left[\begin{array}{c|cc} 5 & \multicolumn{2}{c}{\begin{bmatrix} u & v \end{bmatrix} S} \\ \hline 0 & 2 & w \\ 0 & 0 & 2 \end{array}\right]$$

3. A is normal provided $A^*A = AA^*$.

(a) If A is symmetric, then $A^*A = AA = AA^*$. So A is normal.

(b) If A is real and $A^T = -A$, then $A^*A = (-A)A = A(-A) = AA^*$. So such A is normal.

(c) $(U^*DU)^*U^*DU = U^*D^*DU = U^*DD^*U = (U^*DU)(U^*DU)^*$.

(d) Let $B = AA^*$. The (n,n) entry of AA^* is the sum of the squares of the moduli of the entry in the n column of AA^*. The (n,n)-entry of A^*A is the sum of the squares of the entries in the nth row of A^*A. So $|b_{1n}|^2 + |b_{2n}|^2 + \cdots |b_{nn}|^2 = 0^2 + 0^2 + \cdots + 0^2 + |b_{nn}|^2$. So the only nonzero entry in the last column of B is the diagonal entry.

Now one repeats the similar argument with row $n-1$.

4. (a) Let (λ, x) be an e-pair of A. Consider x^*Ax. On the one hand this is $x^*(Ax) = x^*(\lambda x) = \lambda x^*x$. On the other hand this is $(Ax)^*x = (\lambda x)^*x = \bar{\lambda}x^*x$. So $\lambda x^*x = \bar{\lambda}x^*x$. Since, $x \neq \vec{0}$, we can cancel the x^*x's to get $\lambda = \bar{\lambda}$, that is, λ is real.

(b) We know that $A^* = A$ and $B^* = B$. $(A+B)^{ast} = A^* + B^* = A + B$. So $A + B$ is hermitian.

(c) $(UDU^*)^* = (U^*)^*D^*U^* = UD^*U^{ast} = UDU^*$.

(d) Let (λ, x) be an eigenpair. Use Gram-Schmidt to find a unitary matrix U whose first column is a multiple of x. The the $(j,1)$, entry of U^*AU is λ times the dot-product between column j and 1 of U. So the first column of U^*AU is $\begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. One can check that since A is hermitian, so is U^*AU. So U^*AU has the form

$$\left[\begin{array}{c|ccc} \lambda & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right],$$

where B is a hermitian matrix. Now one continues the same process with B.

5. (a) Let (λ, x) be an eigenpair. Since A is hermitian, λ is real. Now $x^*Ax = x^*(\lambda x) = \lambda x^*x = \lambda\|x\|^2$. Since $x^*Ax \geq 0$ and $\|x\|^2 > 0$, we conclude that $\lambda > 0$.

(b) Let D be a positive matrix. Then its eigenvalues are its diagonal entries. So by (a), if D is positive definite, each of its diagonal entries is necessarily a positive real number.

Now suppose that each diagonal entry of D is positive. Then $x^*Dx = \sum_{j=1}^{n} d_j|x_j|^2$. This is clearly nonnegative, and can only be zero if $x = \vec{0}$. So D is positive definite.

(c) Let $y = Ux$. As x runs over all vectors, so does y, and $x = \vec{0}$ if and only if $y = \vec{0}$. Now $y^*Ay = x^*U^*AUx$, so $y^*Ay > 0$ for all $y \neq 0$ if and only if $x^*(U^*AU)x > 0$ for all $x \neq 0$. Thus, A is positive definite if and only if U^*AU is positive definite.

(d) By 4(d), there exists a unitary U, and a real diagonal matrix D such that $U*AU = D$. By (c), A is positive definite if and only if D is. By 5(b), D is positive definite if and only if each of its diagonal entries is positive.

6. (a) For eigenvalue 6 the vector $u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ spans the eigenspace. For the eigenvalue 3, the vectors $u_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and $u_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ span the eigenspace. We need to make these into an orthonormal basis. So we use Gram-Schmidt. We get exactly Q to be the R in problem 2.

(b) Since the columns of Q come from eigenvectors of A and $Q^{-1} = Q^T$, we have that $Q^TAQ = \text{diag}(6,2,2)$.

(c) Let E be the matrix obtained from D by taking the square root of each diagonal entry. Then $EE^T = D$. So we can take $C = QEQ^T$. Then $CC^T = QEQ^T(QEQ^T)^T = QEQ^TQE^TQ^t = QEE^TQ^T = QDQ^T = A$.

(d) The positive definite square root of A is simply QEQ^T.

(e) Sure. Let F be the matrix obtained from D by taking the 6th root of each diagonal entry. then QFQ^T is a positive definite 6th root of A.

7. Suppose that (λ, x) is an eigenpair of the unitary matrix U. Then $Ux = \lambda x$. Since U is unitary, x and Ux have the same length. So λx and x have the same length. This requires that $|\lambda| = 1$.

8. (a) $x^T Dx = -2x_1^2 - x_2^2 + 3x_4^2$

   (b) The largest would be if we set $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, and $x_4 = 1$. This gives $x^T DX = 3$.

   (c) Since A is symmetric there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^T AQ = D$. Now $x^T Q^T AQx = x^T Dx$. Letting $y = Qx$, we see that as x runs over all vectors of length 1, y will run over all vectors of length 1, and $y^T Ay = x^T Dx$. Thus, the max of $y^T Ay$ over all y of length 1 is the same as the max of $x^T Dx$ over all x of length 1 By b, this maximum is 3

   (d) The only way $x^T Dx = 1$ and $x^T x = 1$ is if $x_1 = x_2 = x_3 = 0$, and $x_4 = \pm 1$ Thus, the only way $y^T Dy = 1$ is if y is $\pm$last column of Q, which is an eigenvector of A corresponding to the eigenvalue 3.