CAAM 453/553 · NUMERICAL ANALYSIS I

Problem Set 1 · Solutions

Posted Wednesday 26 August 2009. Due Friday 4 September 2009.
CAAM 453 students should complete problems 1, 2, 4, 5 [100 points].
CAAM 553 students should complete problems 1, 2, 3, 5, 6, 7. [150 points].
(Students are welcome to attempt more problems if they wish.)

1. [20 points: 10 points per part]

   Recall that λ is an eigenvalue of $\mathbf{A}$ if there exists some vector $\mathbf{v} \neq \mathbf{0}$ such that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$.

   Let $\rho(\mathbf{A})$ denote the *spectral radius* of $\mathbf{A}$, i.e.,

   $$\rho(\mathbf{A}) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{A}\}.$$

   (a) Show that for any *induced* matrix norm, $\|\mathbf{A}\| \geq \rho(\mathbf{A})$ for all $\mathbf{A} \in \mathbb{C}^{n\times n}$.
   [Trefethen & Bau, exercise 3.2]

   (b) Show that the spectral radius *is not* a matrix norm.
   Which of the three basic norm axioms (positivity, scaling, triangle inequality) fail to hold?

Solution.

(a) For any eigenvalue λ of $\mathbf{A}$, there exists a vector $\mathbf{v} \neq 0$ such that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Thus

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \geq \frac{\|\mathbf{A}\mathbf{v}\|}{\|\mathbf{v}\|} = \frac{|\lambda|\,\|\mathbf{v}\|}{\|\mathbf{v}\|} = |\lambda|.$$

As this holds for any eigenvalue of $\mathbf{A}$, it holds for the one for which $|\lambda| = \rho(\mathbf{A})$. Thus we conclude that $\|\mathbf{A}\| \geq \rho(\mathbf{A})$. ■

(b) The spectral radius is *not* a matrix norm, because positivity and the triangle inequality fail to hold.

Positivity *fails*: The matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

has $\rho(\mathbf{A}) = 0$ but $\mathbf{A} \neq \mathbf{0}$.

Scaling *holds*: If λ is an eigenvalue of $\mathbf{A}$, then $\alpha\lambda$ is an eigenvalue of $\alpha\mathbf{A}$, with the same eigenvector.

Triangle inequality *fails*: The matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

both have only zero eigenvalues, yet $\mathbf{A}+\mathbf{B}$ has eigenvalues ± 1. Hence $1 = \rho(\mathbf{A}+\mathbf{B}) > \rho(\mathbf{A}) + \rho(\mathbf{B}) > 0$.

---

2. [15 points]

   Suppose $\mathbf{D} \in \mathbb{C}^{n\times n}$ is zero everywhere except for the main diagonal, $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_n)$. Show that $\|\mathbf{D}\|_p = \max_j |d_j|$ for all $p \geq 1$, where $\|\mathbf{D}\|_p$ refers to the matrix norm induced by the vector p-norm

   $$\|\mathbf{x}\|_p = \Big(\sum_{j=1}^{n} |x_j|^p\Big)^{1/p}.$$

Solution. Since for any $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{Dx}$ is a vector whose jth entry is $d_j x_j$, we have

$$\|\mathbf{D}\|_p^p \;=\; \max_{\|\mathbf{x}\|_p=1} \|\mathbf{Dx}\|_p^p \;=\; \max_{\|\mathbf{x}\|_p=1} \sum_{j=1}^n |d_j x_j|^p.$$

Let $m \in \{1, \ldots, n\}$ be such that $|d_m| = \max_{1 \le j \le n} |d_j|$.

$$\|\mathbf{D}\|_p^p \;=\; \max_{\|\mathbf{x}\|_p=1} \sum_{j=1}^n |d_j x_j|^p \;\le\; |d_m|^p \max_{\|\mathbf{x}\|_p=1} \sum_{j=1}^n |x_j|^p \;=\; |d_m|^p,$$

and so $\|\mathbf{D}\|_p \le |d_m|$. However, note that if we take $\mathbf{x}$ to be zero everywhere except for $x_m = 1$ ($x_j = 0$ if $j \ne m$), then $\|\mathbf{x}\|_p = 1$ and

$$\|\mathbf{Dx}\|_p = |d_m|,$$

and hence $\|\mathbf{D}\|_p = |d_m|$. ■

---

3. [20 points: 5 points per part] Let $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{A} \in \mathbb{C}^{m \times n}$. Prove the following relationships between the 1- and 2-vector norms, and the matrix norms they induce. In each case, demonstrate a vector or matrix for which equality is satisfied.

   (a) $\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1$.

   (b) $\|\mathbf{x}\|_1 \le \sqrt{n}\,\|\mathbf{x}\|_2$.

   (c) $\|\mathbf{A}\|_1 \le \sqrt{m}\,\|\mathbf{A}\|_2$.

   (d) $\|\mathbf{A}\|_2 \le \sqrt{n}\,\|\mathbf{A}\|_1$.

Solution.

(a) We can prove this claim as follows:

$$\|\mathbf{x}\|_1^2 \;=\; \Big(\sum_{j=1}^n |x_j|\Big)^2 \;=\; \sum_{j,k=1}^n |x_j||x_k| \;=\; \sum_{j=1}^n |x_j|^2 + \sum_{j,k=1,j\ne k}^n |x_j||x_k| \;\ge\; \sum_{j=1}^n |x_j|^2 \;=\; \|\mathbf{x}\|_2^2.$$

Equality is realized, for example, if $x_1 = 1$ and $x_j = 0$ for $2 \le j \le n$. ■

(b) The statement $\|\mathbf{x}\|_1 \le \sqrt{n}\|\mathbf{x}\|_2$ is obviously equivalent to $\|\mathbf{x}\|_1^2 \le n\|\mathbf{x}\|_2^2$, which in turn holds if only if

$$\Big(\sum_{j=1}^n |x_j|\Big)^2 \;\le\; n\Big(\sum_{j=1}^n |x_j|^2\Big).$$

We can expand the left hand side to obtain the equivalent expression

$$\sum_{j=1}^n |x_j|^2 + \sum_{j,k=1,j\ne k}^n |x_j||x_k| \;\le\; n\Big(\sum_{j=1}^n |x_j|^2\Big),$$

which can also be written as

$$\sum_{j=1}^n |x_j|^2 + \sum_{j=1}^n \sum_{k=j+1}^n 2|x_j||x_k| \;\le\; n\Big(\sum_{j=1}^n |x_j|^2\Big),$$

Hence we need to prove that

$$0 \;\le\; (n-1)\Big(\sum_{j=1}^{n} |x_j|^2\Big) - \sum_{j=1}^{n}\sum_{k=j+1}^{n} 2|x_j||x_k|.$$

However, the quantity on the right can be rearranged to yield

$$(n-1)\Big(\sum_{j=1}^{n} |x_j|^2\Big) - \sum_{j=1}^{n}\sum_{k=j+1}^{n} 2|x_j||x_k| = \sum_{j=1}^{n}\sum_{k=j+1}^{n} (|x_j|-|x_k|)^2,$$

which, as the sum of squares, is clearly greater than or equal to zero.

Equality is realized when all the entries of $\mathbf{x}$ are identical. ■

At least one student provided a slicker proof based on the Cauchy–Schwarz inequality

$$|\mathbf{y}^*\mathbf{z}| \le \|\mathbf{y}\|_2\|\mathbf{z}\|_2$$

with $\mathbf{y} = (1,1,\ldots,1)^T$ and $\mathbf{z} = |\mathbf{x}|$.

(c) Note that for any $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{Ax} \in \mathbb{C}^m$. Thus, part (b) implies that $\|\mathbf{Ax}\|_1 \le \sqrt{m}\|\mathbf{x}\|_2$. Part (a) can be flipped to see that

$$\frac{1}{\|\mathbf{x}\|_1} \le \frac{1}{\|\mathbf{x}\|_2}.$$

From these two inequalities, it follows that

$$\|\mathbf{A}\|_1 \;=\; \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1} \;\le\; \max_{\mathbf{x}\neq\mathbf{0}} \frac{\sqrt{m}\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \;=\; \sqrt{m}\|\mathbf{A}\|_2.$$

Equality is realized for the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

We see that $\|\mathbf{A}\|_1 = 2$ (the maximum absolute column sum). Recall that $\|\mathbf{A}\|_2$ is the square root of the largest eigenvalue of $\mathbf{A}^*\mathbf{A}$. In this case,

$$\mathbf{A}^*\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and the eigenvalues of this matrix are 0 and 2. Hence we see that $\|\mathbf{A}\|_2 = \sqrt{2}$. Thus,

$$\|\mathbf{A}\|_1 = 2 = \sqrt{2}\cdots\sqrt{2} = \sqrt{2}\|\mathbf{A}\|_2$$

as required. (This example can be generalized to arbitrary dimension by letting $\mathbf{A}$ be zero everywhere except for ones in the first row.) ■

(d) This solution is much like part (c). Part (a) above shows that $\|\mathbf{Ax}\|_2 \le \|\mathbf{Ax}\|_1$, while part (b) shows that

$$\frac{1}{\|\mathbf{x}\|_2} \le \frac{\sqrt{n}}{\|\mathbf{x}\|_1}.$$

Together these imply that

$$\|\mathbf{A}\|_2 \;=\; \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \;\le\; \max_{\mathbf{x}\neq\mathbf{0}} \frac{\sqrt{n}\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1} \;=\; \sqrt{n}\|\mathbf{A}\|_1.$$

Equality is realized for

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

It is clear that $\|\mathbf{A}\|_1 = 1$. To compute $\|\mathbf{A}\|_2$, we see that

$$\mathbf{A}^*\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

which again has eigenvalues 0 and 2. Thus the square root of the largest eigenvalue of $\mathbf{A}^*\mathbf{A}$ is $\sqrt{2}$. Hence $\|\mathbf{A}\|_2 = \sqrt{2} = \sqrt{2}\|\mathbf{A}\|_1$. (This matrix can be generalized to arbitrary dimension by letting $\mathbf{A}$ be zero everywhere except for ones in the first column.) ■

---

4. [20 points: 5 points per part]

   Consider the vector $\mathbf{x} = (2,\, 3/2)^T$.

   (a) Compute a vector $\mathbf{v}$ such that the Householder reflector $\mathbf{H}(\mathbf{v})$ yields

   $$\mathbf{H}(\mathbf{v})\mathbf{x} = \begin{bmatrix} \|\mathbf{x}\|_2 \\ \mathbf{0} \end{bmatrix}.$$

   (b) Compute $\mathbf{H}(\mathbf{v})^*\mathbf{H}(\mathbf{v})$ for this particular $\mathbf{v}$ to verify that $\mathbf{H}(\mathbf{v})$ is unitary.

   (c) Construct the orthogonal projector $\mathbf{P}$ onto $\text{span}\{\mathbf{v}\}$.

   (d) Produce a precise drawing (or plot) showing $\text{span}\{\mathbf{v}\}$, $\text{span}\{\mathbf{v}\}^\perp$, $\mathbf{x}$, $\mathbf{P}\mathbf{x}$, $(\mathbf{I}-\mathbf{P})\mathbf{x}$, and $\mathbf{H}(\mathbf{v})\mathbf{x}$. Be sure to label your illustration clearly.

Solution.

(a) Note that $\|\mathbf{x}\|_2 = \sqrt{2^2 + (3/2)^2} = 5/2$. Set $\mathbf{v} = \mathbf{x} - \|\mathbf{x}\|\mathbf{e}_1 = (2,\, 3/2)^T - (5/2,\, 0)^T = (-1/2,\, 3/2)$, giving

$$\mathbf{H}(\mathbf{v}) = \mathbf{I} - 2\frac{\mathbf{v}\mathbf{v}^*}{\mathbf{v}^*\mathbf{v}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \tfrac{2}{5/2}\begin{pmatrix} 1/4 & -3/4 \\ -3/4 & 9/4 \end{pmatrix} = \begin{pmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{pmatrix}.$$

We can verify that

$$\mathbf{H}(\mathbf{v})\mathbf{x} = \begin{pmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{pmatrix}\begin{pmatrix} 2 \\ 3/2 \end{pmatrix} = \begin{pmatrix} 16/10 + 9/10 \\ 12/10 - 12/10 \end{pmatrix} = \begin{pmatrix} 5/2 \\ 0 \end{pmatrix}.$$

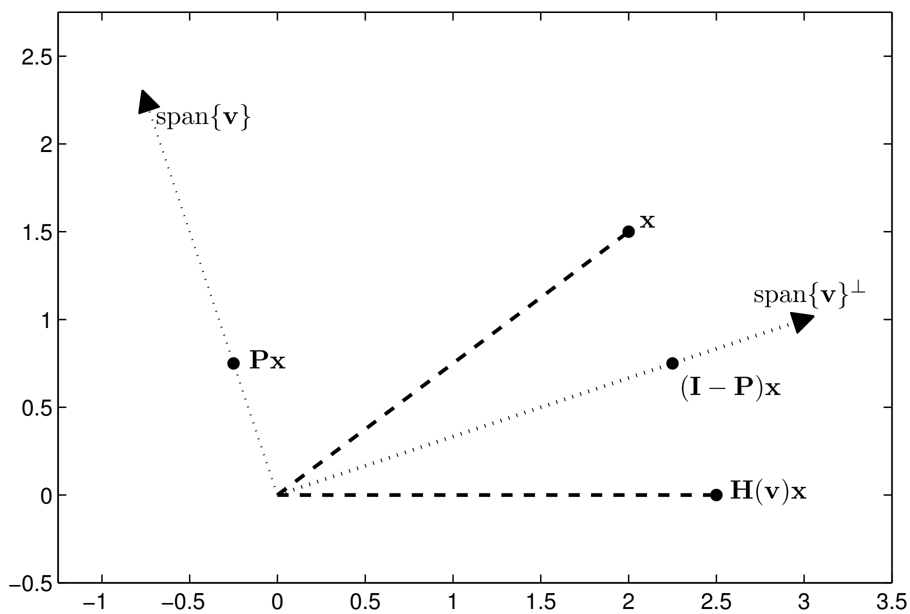
(b)

$$\mathbf{H}(\mathbf{v})^*\mathbf{H}(\mathbf{v}) = \begin{pmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{pmatrix}\begin{pmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{pmatrix} = \begin{pmatrix} 16/25 + 9/25 & 12/25 - 12/25 \\ 12/25 - 12/25 & 9/25 + 16/25 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(c)

$$\mathbf{P} = \frac{\mathbf{v}\mathbf{v}^*}{\mathbf{v}^*\mathbf{v}} = \tfrac{1}{5/2}\begin{pmatrix} 1/4 & -3/4 \\ -3/4 & 9/4 \end{pmatrix} = \begin{pmatrix} 1/10 & -3/10 \\ -3/10 & 9/10 \end{pmatrix}.$$

(d)



---

5. (a)=5 points, (b)=5 points, (c)=15 points, (d)=20 points]

For a dense matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the standard technique for computing the QR decomposition uses Householder reflectors. An alternative approach reduces $\mathbf{A}$ to upper-triangular form using *Givens rotations*. A Givens rotation is a unitary matrix whose entries are zero except as specified:

$$\mathbf{G}(j,k,\theta) = \begin{bmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & \cos\theta & & & & \sin\theta & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & -\sin\theta & & & & \cos\theta & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow \textit{row } j \\ \\ \\ \\ \leftarrow \textit{row } k \\ \\ \\ \\ \end{matrix}.$$

$$\qquad\qquad\qquad\uparrow \textit{column } j \qquad\qquad \uparrow \textit{column } k$$

(a) Prove that the 2×2 Givens rotation

$$\mathbf{G}(1,2,\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

is a unitary matrix for any θ.

(b) Show how to pick θ such that

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sqrt{\alpha^2 + \beta^2} \\ 0 \end{bmatrix}.$$

(c) Design an algorithm for computing the QR factorization $\mathbf{A} = \mathbf{QR}$ using Givens rotations, rather than Householder transformations. In particular, describe the order in which the lower triangular entries of $\mathbf{A}$ (that is, the entries in positions (j, k) for $j > k$) should be eliminated, and those Givens rotations that should be used to achieve this elimination. How many total rotations are necessary? What is the matrix $\mathbf{Q}$?

(d) Write a MATLAB function

```
function [Q,R] = givens_qr(A)
```

that constructs the matrices $\mathbf{Q}$ and $\mathbf{R}$ for $\mathbf{A} \in \mathbb{C}^{m\times n}$ using the algorithm you describe in part (c). Demonstrate that your routine works by reporting `norm(Q*R-A)/norm(A)` for `A = randn(10,5)` and the *final* matrix `R`. *Please do not print out* `Q` *or the values of* `R` *at each step.*

Solution.

(a) Proof. We can show $\mathbf{G}(1, 2, \theta)$ is a unitary matrix via a simple computation:

$$\begin{aligned}
\mathbf{G}(1,2,\theta)^*\mathbf{G}(1,2,\theta) &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\
&= \begin{pmatrix} \cos^2\theta + \sin^2\theta & \cos\theta\sin\theta - \sin\theta\cos\theta \\ \sin\theta\cos\theta - \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad \blacksquare
\end{aligned}$$

(b) The equation

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sqrt{\alpha^2+\beta^2} \\ 0 \end{pmatrix}$$

can be broken into the requirements

$$\begin{aligned}
\alpha\cos\theta + \beta\sin\theta &= \sqrt{\alpha^2+\beta^2} \\
\beta\cos\theta - \alpha\sin\theta &= 0.
\end{aligned}$$

The first equation will immediately be satisfied if the second equation is satisfied, as the 2-norm is preserved by unitary transformations, and $\|(\alpha, \beta)^T\|_2 = \sqrt{\alpha^2+\beta^2}$. Thus, we need only pick θ such that

$$\beta\cos\theta = \alpha\sin\theta.$$

This is equivalent to $\tan\theta = \beta/\alpha$, i.e.,

$$\theta = \tan^{-1}(\beta/\alpha).$$

(c) To compute a QR factorization of $\mathbf{A}$ using Givens rotations, we must successively zero out entries below the main diagonal of $\mathbf{A}$. Premultiplying $\mathbf{A}$ by $\mathbf{G}(1, 2, \theta_{2,1})$ with $\theta_{2,1} = \tan^{-1}(a_{2,1}/a_{1,1})$ will introduce a zero into the $(2, 1)$ entry of A. Note that this transformation effectively replaces the first two rows of $\mathbf{A}$ with a linear combination of each other, but does not affect any other entries in $\mathbf{A}$ outside the first two rows.

Next, introduce a zero into the $(3, 1)$ position by premultiplying $\mathbf{G}(1, 2, \theta_{2,1})\mathbf{A}$ by $\mathbf{G}(1, 3, \theta_{3,1})$ for appropriate choice of $\theta_{3,1}$. Key observation: this will not overwrite the zero we have already created in the $(2,1)$ position! (If we had tried to use $\mathbf{G}(2,3,\theta)$ to zero out the $(3,1)$ position, we would have overwritten that zero in $(2,1)$.)

One proceeds in this fashion to introduce zeros down the first column. Then, on premultiplies by $\mathbf{G}(2,3,\theta_{3,2})$ to eliminate the $(3,2)$ entry, the first subdiagonal entry in the second column. Eliminate the rest of the second column, move to the third, and so on. In the end, one has

$$\mathbf{G}(n,m,\theta_{m,n})\cdots\mathbf{G}(n,n+1,\theta_{n+1,n})\mathbf{G}(n-1,m,\theta_{m,n-1})\cdots\mathbf{G}(1,m,\theta_{m,1})\cdots\mathbf{G}(1,2,\theta_{2,1})\mathbf{A}=\mathbf{R}.$$

This is more easily explained in MATLAB than in words! See problem 6(d) below.

Overall, we have applied $m-1$ rotations to put zeros in the first column, $m-2$ rotations for the second column, etc., with $m-n$ rotations to zero the subdiagonal entries in the nth column. Thus our algorithm requires

$$\sum_{j=1}^{n}(m-j)=mn-\tfrac{1}{2}n(n+1)$$

total rotations.

The matrix $\mathbf{Q}$ is obtained by repeatedly premultiplying the above equation by matrices of the form $\mathbf{G}(j,k,\theta)^*$:

$$\mathbf{Q}=\mathbf{G}(1,2,\theta_{2,1})^*\mathbf{G}(1,m,\theta_{m,1})^*\cdots\mathbf{G}(n,m,\theta_{m,n})^*.$$

As an aside, one might wonder whether to use Householder transformations or Givens rotations to compute the QR decomposition. In efficient implementations, it turns out that the Givens QR algorithm requires roughly 1.5 times as many floating point operations as the Householder QR algorithm, so Householder is favored. However, if the matrix $\mathbf{A}$ already has many zero entries in its lower triangle (as often occurs in applications), then the Givens QR algorithm can exploit this sparsity and compute a QR factorization more quickly than the Householder QR algorithm.

(d) Here is a basic, inefficient MATLAB implementation of the Givens QR algorithm.

```
function [Q,R] = givens_qr(A)
if ~isreal(A), fprintf('Error: A must be a matrix with real entries.\n'), break, end
[m,n] = size(A);
Q = eye(m);
for j=1:n
   for k=j+1:m
      G = eye(m);
      theta = atan(A(k,j)/A(j,j));       % Form theta to zero out A(k,j)
      G(j,j) =  cos(theta);
      G(k,k) =  cos(theta);
      G(j,k) =  sin(theta);
      G(k,j) = -sin(theta);              % Form Givens rotation G(j,k,theta)
      A = G*A;                           % Update A
      Q = Q*G';                          % Form Q
   end
end
R = A;
```

Here is a slicker method that avoids the large matrix-matrix products and has other numerical advantages.

```
function [Q,R] = slick_givens_qr(A)
[m,n] = size(A);
Q = eye(m);
for j=1:n
   for k=j+1:m
      if abs(A(k,j)) == 0                   % We use a numerically superior technique
            c = 1; s = 0;                   % for computing c = cos(theta) and s = sin(theta).
      elseif abs(A(k,j)) > abs(A(j,j))      % See Golub & Van Loan, Matrix Computations, Algorithm 5.1.5
```

```
                tau = A(j,j)/A(k,j);
                s = 1/sqrt(1+tau^2);
                c = s*tau;
          else
                tau = A(k,j)/A(j,j);
                c = 1/sqrt(1+tau^2);
                s = c*tau;
          end
          A([j k],:) = [c s; -s c]*A([j k],:);    % G(j,k,theta) only affects rows j & k of A
          Q(:,[j k]) = Q(:,[j k])*[c -s; s c];    % Q*G(j,k,theta)^* only affects columns j & k of Q
       end
    end
    R = A;
```

Here is are MATLAB results showing `givens_qr` in action.

```
>> A = randn(10,5)
A =
   -1.1859    0.1286    0.8057   -0.3306   -0.1199
   -1.0559    0.6565    0.2316   -0.8436   -0.0653
    1.4725   -1.1678   -0.9898    0.4978    0.4853
    0.0557   -0.4606    1.3396    1.4885   -0.5955
   -1.2173   -0.2624    0.2895   -0.5465   -0.1497
   -0.0412   -1.2132    1.4789   -0.8468   -0.4348
   -1.1283   -1.3194    1.1380   -0.2463   -0.0793
   -1.3493    0.9312   -0.6841    0.6630    1.5352
   -0.2611    0.0112   -1.2919   -0.8542   -0.6065
    0.9535   -0.6451   -0.0729   -1.2013   -1.3474
>> [Q,R] = givens_qr(A)
Q =
    0.3705   -0.0711   -0.1405    0.0911    0.0566    0.2037   -0.7126   -0.5216    0.0452    0.0554
    0.3299    0.1593   -0.0731    0.3390   -0.0523   -0.8559    0.0146   -0.0767   -0.0544   -0.0330
   -0.4600   -0.3258    0.4552   -0.3045   -0.2510   -0.3760   -0.4017   -0.0889   -0.0902    0.0212
   -0.0174   -0.1834   -0.4470   -0.5328    0.5666   -0.2803   -0.0980    0.1540    0.1957    0.1055
    0.3803   -0.2351    0.1947    0.0626    0.1588    0.0759   -0.2586    0.6141   -0.5323   -0.0057
    0.0129   -0.5027   -0.2735    0.2907   -0.3737    0.0104   -0.1203    0.3601    0.5308   -0.1489
    0.3525   -0.6600    0.1175   -0.1295   -0.0165         0    0.4896   -0.3862   -0.1182    0.0806
    0.4215    0.2415    0.2947   -0.3481   -0.2874         0         0    0.1857    0.3841    0.5397
    0.0816   -0.0227    0.5950    0.1879    0.5357         0         0         0    0.4712   -0.3073
   -0.2979   -0.1654    0.0233    0.4883    0.2742         0         0         0         0    0.7549
R =
   -3.2011    0.8145    0.9651   -0.3936    0.6296
   -0.0000    2.4342   -1.6300    0.5488    0.8626
    0.0000    0.0000   -2.3658   -0.5751    0.6494
   -0.0000   -0.0000   -0.0000   -2.4872   -1.2953
   -0.0000   -0.0000   -0.0000    0.0000   -1.4582
   -0.0000    0.0000    0.0000    0.0000   -0.0000
   -0.0000    0.0000   -0.0000    0.0000    0.0000
   -0.0000   -0.0000   -0.0000    0.0000    0.0000
   -0.0000    0.0000   -0.0000   -0.0000   -0.0000
    0.0000    0.0000    0.0000    0.0000   -0.0000

>> norm(Q*R-A)/norm(A)
ans =
   3.0202e-16
```

We confirm that, for this random example, `givens_qr` generates a correct result. (Error on the order of 10^{-16} is expected from computer roundoff errors.)

---

6. [25 points: 5 points per part]

   Let $|\mathbf{x}|$ denote the entrywise absolute value of a vector $\mathbf{x}$, i.e., if x_j is the jth entry of $\mathbf{x}$, then the jth entry of $|\mathbf{x}|$ is $|x_j|$. We say that $|\mathbf{x}| \le |\mathbf{y}|$ provided $|x_j| \le |y_j|$ for all j.

   A vector norm $\|\cdot\|$ is said to be *absolute* provided $\|\mathbf{x}\| = \|\,|\mathbf{x}|\,\|$.

   A vector norm $\|\cdot\|$ is said to be *monotone* if $|\mathbf{x}| \le |\mathbf{y}|$ implies $\|\mathbf{x}\| \le \|\mathbf{y}\|$.

Consider the following vector norm on $\mathbb{R}^2$:

$$\|\mathbf{x}\| = |x_1 - x_2| + |x_2|,$$

where $\mathbf{x} = (x_1, x_2)^T$.

(a) Show that this norm satisfies the three vector norm axioms.

(b) Draw the unit ball for this norm (i.e., the set of all $\mathbf{x} \in \mathbb{R}^2$ for which $\|x\| = 1$).

(c) Is this norm absolute?

(d) Is this norm monotone?

(e) Compute the matrix norm induced by this vector norm for the matrix

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

[adapted from Horn and Johnson]

Solution.

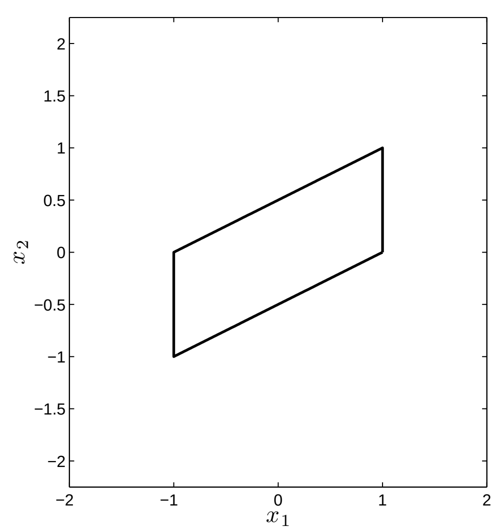
(a) *Positivity*: By definition it is evident that $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in \mathbb{R}^2$. If $\mathbf{x} = \mathbf{0}$, then $|x_1 - x_2| + |x_2| = 0$, which of course implies $|x_1 - x_2| = 0$ and $|x_2| = 0$: Thus $x_1 = x_2$ and $x_2 = 0$. Hence, $\|\mathbf{x}\| = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$.
Scaling: $\|\alpha\mathbf{x}\| = |\alpha x_1 - \alpha x_2| + |\alpha x_2| = |\alpha||x_1 - x_2| + |\alpha||x_2| = |\alpha|\|\mathbf{x}\|$.
Triangle inequality: $\|\mathbf{x} + \mathbf{y}\| = |(x_1 + y_1) - (x_2 + y_2)| + |x_2 + y_2|$. The triangle inequality for real numbers implies

$$\|\mathbf{x} + \mathbf{y}\| = |(x_1 + y_1) - (x_2 + y_2)| + |x_2 + y_2| \leq |x_1 - x_2| + |y_1 - y_2| + |x_2| + |y_2| = \|\mathbf{x}\| + \|\mathbf{y}\|.$$

(b) One can determine all the points $\mathbf{x} \in \mathbb{R}^2$ for which $\|\mathbf{x}\| = 1$ by considering all cases of the absolute values individually; e.g., $x_1 - x_2 \geq 0$ and $x_2 \geq 0$ implies that $\|\mathbf{x}\| = x_1 - x_2 + x_2 = x_1$, so the only values of $\mathbf{x}$ having such coordinates and $\|\mathbf{x}\| = 1$ are:

$$\left\{ \begin{bmatrix} 1 \\ x_2 \end{bmatrix} : x_2 \in [0, 1] \right\}.$$

Repeat this for the three remaining possibilities ($x_1 - x_2 \leq 0$ and $x_2 \geq 0$; $x_1 - x_2 \leq 0$ and $x_2 \leq 0$; $x_1 - x_2 \geq 0$ and $x_2 \leq 0$) to obtain vectors that trace out the unit ball shown in the plot below.



(c) *No*, this norm *is not absolute.* For example,

$$\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = |1-1| + |1| = 1, \qquad \left\| \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\| = |1+1| + |-1| = 3.$$

(Notice that you tell that the norm is not absolute by the shape of the unit ball.)

(d) *No*, this norm *is not monotone.* For example

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

but

$$\left\| \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\| = 2 > 1 = \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|.$$

(e) To compute the induced norm of the matrix $\mathbf{D}$, we first compute

$$\mathbf{D}\mathbf{x} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 2x_2 \end{bmatrix},$$

and hence $\|\mathbf{D}\mathbf{x}\| = |x_1 - 2x_2| + |2x_2|$. We need to compute the maximum value of this function over all vectors $\mathbf{x}$ on the unit ball computed in part (a). This is a bit of a tedious procedure. For example, the norm of the vertical segment of the map $\mathbf{D}\mathbf{x}$ corresponding to $x_1 = 1$ and $x_2 \in [0,1]$ breaks down into two cases:

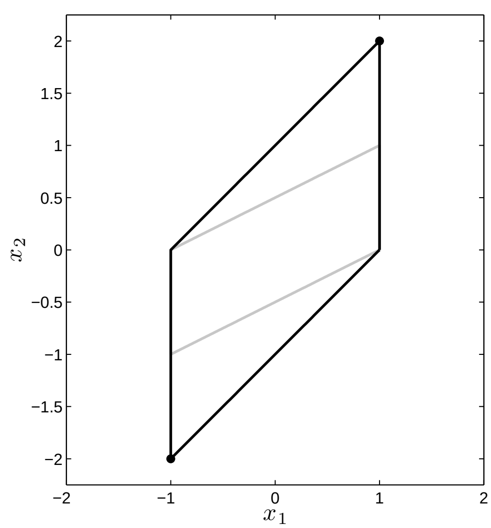
$$\left\| \begin{bmatrix} 1 \\ 2x_2 \end{bmatrix} \right\| = \begin{cases} 1, & x_2 \in [0, 1/2]; \\ 4x_2 - 1, & x_2 \in [1/2, 1]. \end{cases}$$

Hence the value of $\|\mathbf{D}\mathbf{x}\|$ ranges between 1 ($x_2 = 0$) and 3 ($x_2 = 1$) on this segment. Repeating similar calculations for each segment of the unit ball, we find that $\|\mathbf{D}\mathbf{x}\| \in [1,3]$ for all $\mathbf{x}$ on the unit fall, with equality attained at $\mathbf{x} = [1,1]^T$ and $\mathbf{x} = [-1,-1]^T$. That is,

$$\|\mathbf{D}\| = 3.$$

(Notice that for this norm and this $\mathbf{D}$, we have $\|\mathbf{D}\| > \max_k |d_{k,k}|$.

The plot below shows the image of the unit ball under the matrix $\mathbf{D}$ (along with the unit ball, in gray, to facilitate comparison).



7. [25 points: (a)=9 points, (b)=8 points, (c)=8 points]

Projectors play an essential role in spectral theory, that is, the study of eigenvalues, eigenvectors, and related objects. Suppose $\mathbf{A} \in \mathbb{C}^{n\times n}$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ with associated right eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and left eigenvectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ (i.e., $\mathbf{A}\mathbf{x}_j = \lambda_j \mathbf{x}_j$ and $\mathbf{y}_j^*\mathbf{A} = \lambda_j \mathbf{y}_j^*$). Then the *spectral projector* associated with λ_j is defined as

$$\mathbf{P}_j = \frac{\mathbf{x}_j\mathbf{y}_j^*}{\mathbf{y}_j^*\mathbf{x}_j} \in \mathbb{C}^{n\times n}.$$

(a) Compute the spectral projectors for the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & \alpha \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

for arbitrary constant α. Under what circumstances are these orthogonal projectors?

(b) Verify that $\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 = \mathbf{I}$ and $\mathbf{P}_j\mathbf{P}_k = \mathbf{0}$ when $j \neq k$.
(These properties hold in general for spectral projectors.)

(c) How do the 2-norms of the spectral projectors associated with the eigenvalues $\lambda = -1$ and $\lambda = 1$ relate to the angle between the right eigenvectors associated with these eigenvalues as $\alpha \to \infty$? (A qualitative answer is sufficient; you do not need to explicitly compute the 2-norms.)

(d) [optional] For the matrix $\mathbf{A}$ in part (a), confirm that you obtain the same spectral projectors from the formula

$$\mathbf{P}_j = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_j} (z\mathbf{I} - \mathbf{A})^{-1}\, \mathrm{d}z,$$

where Γ_j is a Jordan curve containing λ_j in its interior and all other eigenvalues in its exterior. (For details on this approach, see Section I.5 of Tosio Kato, *Perturbation Theory for Linear Operators*, corrected 2nd ed., Springer, 1980.)

Solution.

(a) First we compute right and left eigenvectors for each eigenvalue:

$$\lambda_1 = -1, \qquad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{y}_1 = \begin{bmatrix} 1 \\ 0 \\ \overline{\alpha}/2 \end{bmatrix}.$$

$$\lambda_2 = 0, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{y}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

$$\lambda_3 = 1, \qquad \mathbf{x}_3 = \begin{bmatrix} -\alpha/2 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{y}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

from which we construct the spectral projectors:

$$\mathbf{P}_1 = \frac{\mathbf{x}_1\mathbf{y}_1^*}{\mathbf{y}_1^*\mathbf{x}_1} = \begin{bmatrix} 1 & 0 & \alpha/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{P}_2 = \frac{\mathbf{x}_2\mathbf{y}_2^*}{\mathbf{y}_2^*\mathbf{x}_2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{P}_3 = \frac{\mathbf{x}_3\mathbf{y}_3^*}{\mathbf{y}_3^*\mathbf{x}_3} = \begin{bmatrix} 0 & 0 & -\alpha/2 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The projectors $\mathbf{P}_1$ and $\mathbf{P}_2$ are orthogonal if an only if $\alpha = 0$; $\mathbf{P}_2$ is an orthogonal projector for all α.

(b) By inspection one can see that $\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 = \mathbf{I}$ and $\mathbf{P}_j\mathbf{P}_k = \mathbf{0}$ for $j \neq k$.

(c) *Please do not subtract points if the student assumes* $\alpha \in \mathbb{R}$.
The 2-norm of $\mathbf{P}_1$ will scale linearly with α as $|\alpha| \to \infty$. More precisely, $\|\mathbf{P}_1\|_2$ is the square root of the largest eigenvalue of $\mathbf{P}_1^*\mathbf{P}_1$. Note that

$$\mathbf{P}_1^*\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & \alpha/2 \\ 0 & 0 & 0 \\ \overline{\alpha}/2 & 0 & |\alpha|^2/4 \end{bmatrix},$$

whose eigenvalues are 0, 0, and $1 + |\alpha|^2/4$. Hence,

$$\|\mathbf{P}_1\|_2 = \sqrt{1 + |\alpha|^2/4} = \frac{|\alpha|}{2} + O(1/|\alpha|)$$

as $|\alpha| \to \infty$. The angle between $\mathbf{x}_1$ and $\mathbf{y}_1$ is given by

$$\cos\angle(\mathbf{x}_1, \mathbf{y}_1) = \frac{|\mathbf{y}_1^*\mathbf{x}_1|}{\|\mathbf{x}_1\|_2\|\mathbf{y}_1\|_2} = \frac{1}{\sqrt{1 + |\alpha|^2/4}}.$$

Since $\cos\angle(\mathbf{x}_1, \mathbf{y}_1) \to 0$ as $\alpha \to \infty$, we have $\angle(\mathbf{x}_1, \mathbf{y}_1) \to \pi/2$.
The same relationship holds for $\mathbf{P}_3$ and $\angle(\mathbf{x}_3, \mathbf{y}_3)$.
Of course, since $\mathbf{x}_2$ and $\mathbf{y}_2$ are independent of α, both $\|\mathbf{P}_2\|_2$ and $\angle(\mathbf{x}_2, \mathbf{y}_2)$ are independent of α.
The point? There is a very close relationship between the angle between left and right eigenvectors, and the norm of the spectral projector. This has significant implications for eigenvalue perturbation theory.

(d) We will use two important results about complex integrals. Suppose f is an analytic function on the Jordan curve Γ and its interior, and a is contained in that interior. The we have Cauchy's Theorem:

$$\int_\Gamma f(z)\,\mathrm{d}z = 0$$

and the Cauchy integral formula:

$$f(a) = \frac{1}{2\pi\mathrm{i}}\int_\Gamma \frac{f(z)}{z-a}\mathrm{d}z.$$

For the problem at hand, note that when z is not an eigenvalue, we have

$$(z\mathbf{I} - \mathbf{A})^{-1} = \begin{bmatrix} (z+1)^{-1} & 0 & -\alpha(z+1)^{-1}(z-1)^{-1} \\ 0 & z^{-1} & 0 \\ 0 & 0 & (z-1)^{-1} \end{bmatrix}.$$

We consider only $j = 1$; the other cases are similar. Let Γ_1 be a small circle enclosing $\lambda_1 = -1$. The integral of the matrix is the matrix of integrals:

$$\frac{1}{2\pi\mathrm{i}}\int_{\Gamma_1}(z\mathbf{I} - \mathbf{A})^{-1}\,\mathrm{d}z = \frac{1}{2\pi\mathrm{i}}\begin{bmatrix} \int_{\Gamma_1}(z+1)^{-1}\,\mathrm{d}z & \int_{\Gamma_1}0\,\mathrm{d}z & \int_{\Gamma_1}-\alpha(z+1)^{-1}(z-1)^{-1}\,\mathrm{d}z \\ \int_{\Gamma_1}0\,\mathrm{d}z & \int_{\Gamma_1}z^{-1}\,\mathrm{d}z & \int_{\Gamma_1}0\,\mathrm{d}z \\ \int_{\Gamma_1}0\,\mathrm{d}z & \int_{\Gamma_1}0\,\mathrm{d}z & \int_{\Gamma_1}(z-1)^{-1}\,\mathrm{d}z \end{bmatrix}.$$

When $f(z) = 0$, $f(z) = z^{-1}$, and $f(z) = (z-1)^{-1}$, by Cauchy's Theorem we have $\int_{\Gamma_1} f(z)\,\mathrm{d}z = 0$. The only interesting integrals can then be computed using the Cauchy integral formula:

$$\int_{\Gamma_1}\frac{1}{z+1}\,\mathrm{d}z = 2\pi\mathrm{i}$$

and

$$\int_{\Gamma_1} \frac{-\alpha(z-1)^{-1}}{z+1} \mathrm{d}z = 2\pi\mathrm{i}\Big[-\alpha(z-1)^{-1}\Big]_{z=-1} = 2\pi\mathrm{i}\big(\alpha/2\big).$$

Inserting these results into the matrix of integrals yields

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & \alpha/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

in agreement with our result from part (a).

Supplemental Problem

This optional extra problem goes beyond the scope of the lectures.

S1. Let $\mathbf{P} \in \mathbb{C}^{n\times n}$ be any (not necessarily orthogonal) projector with $1 \le \operatorname{rank} \mathbf{P} < n$. Prove that $\|\mathbf{P}\|_2 = \|\mathbf{I}-\mathbf{P}\|_2$.

Hint. This handy theorem has been repeatedly discovered over the years, as described in a very interesting recent survey by Daniel Szyld. Here is an outline of the simplest proof that Szyld provides.

- Explain why one can reduce the problem to proving that $\|\mathbf{Pz}\|_2 \le \|\mathbf{I}-\mathbf{P}\|_2$ for all $\|\mathbf{z}\|_2 = 1$.

To prove this fact, take any $\mathbf{z} \in \mathbb{C}^n$ with $\|\mathbf{z}\|_2 = 1$, and define $\mathbf{x} = \mathbf{Pz}$ and $\mathbf{y} = (\mathbf{I}-\mathbf{P})\mathbf{z}$.

- Explain why $\|\mathbf{Pz}\|_2 \le \|\mathbf{I}-\mathbf{P}\|_2$ if $\mathbf{x} = 0$ or $\mathbf{y} = 0$. (Show that $\|\mathbf{I}-\mathbf{P}\|_2 \ge 1$.)

If $\mathbf{x}$ and $\mathbf{y}$ are both nonzero, then define

$$\widehat{\mathbf{x}} = \frac{\|\mathbf{y}\|_2}{\|\mathbf{x}\|_2}\mathbf{x}, \qquad \widehat{\mathbf{y}} = \frac{\|\mathbf{x}\|_2}{\|\mathbf{y}\|_2}\mathbf{y}, \qquad \mathbf{w} = \widehat{\mathbf{x}} + \widehat{\mathbf{y}}.$$

- Show $\|\mathbf{w}\|_2 = \|\mathbf{z}\|_2$.
- Explain why $\|\mathbf{Pz}\|_2 = \|(\mathbf{I}-\mathbf{P})\mathbf{w}\|_2 \le \|\mathbf{I}-\mathbf{P}\|_2$.

Solution. See Daniel B. Szyld, "The many proofs of an identity on the norm of oblique projections," *Numerical Algorithms* 42 (2006) 309–323.
