CMSC 666
Spring 2007

Homework 4

Roman Stanchak
April 17, 2007

Problem 1: Show the Gram-Schmidt method can be viewed as a technique for computing a QR factorization of A.

Solution 1: The Gram-Schmidt method produces a set of orthonormal basis vectors $\{q_1, q_2, \ldots, q_n\}$. Any arbitrary vector can be represented in this basis as a multiplication of the matrix $Q = [q_1, q_2, \ldots q_n]$ by some vector of coefficients, r. Consider the columns of $A = [a_1, a_2, \ldots a_n]$. The vector a_1 is orthogonalized by normalizing it, that is, $a_1 = ||a_1|| \cdot q_1$, or in other words:

$$\begin{aligned} r_1 &= \begin{bmatrix} ||a_1|| \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \\ &\text{and} \\ Qr_1 &= a_1 \end{aligned}$$

An arbitrary column a_i of A, is orthogonalized by

$$\begin{aligned} \hat{q}_i &= a_i - \sum_{j=1}^{i-1} < a_i, q_j > q_j \\ q_i &= \frac{\hat{q}_i}{||\hat{q}_i||} \\ a_i &= ||\hat{q}_i|| q_i + \sum_{j=1}^{i-1} < a_i, q_j > q_j \end{aligned}$$

Or in vector form

$$\begin{aligned} r_i &= \begin{bmatrix} < a_i, q_1 > \\ < a_i, q_2 > \\ . \\ . \\ . \\ < a_i, q_{i-1} > \\ ||\hat{q}_2|| \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \\ &\text{and} \\ Qr_i &= a_i \end{aligned}$$

Concatenating the vectors r_i into matrix form constructs an upper triangular matrix $R = r_1, r_2, \ldots, r_n$, for which it is true that:

$$
\begin{aligned}
Q \begin{bmatrix} | & | & & | \\ r_1 & r_2 & \dots & r_n \\ | & | & & | \end{bmatrix} &= \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \end{bmatrix} \\
&\text{and} \\
QR &= A
\end{aligned}
$$

Since Q is orthonormal and R is upper triangular, this is precisely a QR factorization of A.

Problem 2: For symmetric matrix A, show that for the Rayleigh quotient, $\mathcal{R}(x)$, $\max_{x \neq 0} \mathcal{R}(x) = \lambda_{max}(A)$, and $\min_{x \neq 0} \mathcal{R}(x) = \lambda_{min}(A)$.

Solution 2:

$$
\begin{aligned}
\mathcal{R}(x) &= \frac{< x, Ax >}{< x, x >} \\
&= \frac{x^* Ax}{x^* x}
\end{aligned}
$$

A is symmetric, so A has a full set of eigenvalues and eigenvectors, and can be broken down using the Schur decomposition

$$
\begin{aligned}
A &= U \Lambda U^* \\
\text{where:} & \quad U \text{ is a unitary matrix whose columns are eigenvectors of } A \\
& \quad \Lambda \text{ is a diagonal matrix whose entries are eigenvalues} \\
\mathcal{R}(x) &= \frac{x^* U \Lambda U^* x}{x^* x} \\
&= \frac{x^* U \Lambda U^* x}{x^* U U^* x} \\
\text{let: } w &= U^* x \\
&= \frac{w^* \Lambda w}{w^* w} \\
&= < w, w >^{-1} \sum (\lambda_i w_i^2)
\end{aligned}
$$

Thus:

$$
\begin{aligned}
\min_{x \neq 0} \mathcal{R}(x) &= \lambda_{min} \\
\max_{x \neq 0} \mathcal{R}(x) &= \lambda_{max}
\end{aligned}
$$

Problem 3a: For the Lanczos recurrence applied to a symmetric matrix A, show that the vectors $\{v_j\}$ make up an orthonormal set.

Solution 3a: Proof by induction on j. **Base case** $j = 2$.
Show $< v_1, v_2 >= 0$, $< v_2, v_2 >= 1$.

$< v_2, v_2 >= 1$ by construction of v_2 as a unit vector.

$$\begin{aligned}
\beta_2 v_2 &= Av_1 - \alpha_1 v_1 - 0 \\
\\
< v_1, v_2 > &= v_1^T \frac{Av_1 - \alpha_1 v_1}{\beta_2} \\
&= \beta_2^{-1} \left(v_1^T A v_1 - \alpha_1 v_1^T v_1 \right) \\
&= \beta_2^{-1} \left(v_1^T A v_1 - \alpha_1 \right) \\
&= \beta_2^{-1} \left(v_1^T A v_1 - v_1^T A v_1 \right) \\
&= 0
\end{aligned}$$

Base case $j = 3$.
Show $< v_1, v_3 >= 0$, $< v_2, v_3 >= 0$, $< v_3, v_3 >= 1$.
Again, $< v_3, v_3 >= 1$ by construction of v_3 as a unit vector.

$$\begin{aligned}
\beta_3 v_3 &= Av_2 - \alpha_2 v_2 - \beta_2 v_1 \\
\\
< v_2, v_3 > &= v_2^T \frac{Av_2 - \alpha_2 v_2 - \beta_2 v_1}{\beta_3} \\
&= \beta_3^{-1} \left(v_2^T A v_2 - \alpha_2 v_2^T v_2 - \beta_2 v_2^T v_1 \right) \\
&= \beta_3^{-1} \left(\alpha_2 - \alpha_2 \cdot 1 - 0 \right) \\
&= 0 \\
\\
< v_1, v_3 > &= v_1^T \frac{Av_2 - \alpha_2 v_2 - \beta_2 v_1}{\beta_3} \\
&= \beta_3^{-1} \left(v_1^T A v_2 - \alpha_2 v_1^T v_2 - \beta_2 v_1^T v_1 \right) \\
&= \beta_3^{-1} \left(v_1^T A v_2 - 0 - \beta_2 \right)
\end{aligned}$$

Note that:

$$\begin{aligned}
\beta_2 &= \frac{< v_2, \beta_2 v_2 >}{< v_2, v_2 >} \\
&= < v_2, \beta_2 v_2 > \\
&= v_2^T A v_1 - \alpha_1 v_2^T v_1 \\
&= v_2^T A v_1
\end{aligned}$$

So:

$$< v_1, v_3 > = \beta_3^{-1} \left(v_1^T A v_2 - v_2^T A v_1 \right)$$

And since A is symmetric:

$$\begin{aligned}
&= \beta_3^{-1} \left(v_1^T A v_2 - v_1^T A v_2 \right) \\
&= 0
\end{aligned}$$

Inductive Case Assume $< v_i, v_j >= \delta_{i,j} \forall i, j < k$. Show for k.

$$\beta_k v_k = Av_{k-1} - \alpha_{k-1} v_{k-1} - \beta_{k-1} v_{k-2}$$

Again, by construction of v_k as a unit vector, $< v_k, v_k >= 1$. Case 1, show $< v_{k-1}, v_k >= 0$:

$$\begin{aligned} < v_{k-1}, v_k > &= \beta_k^- 1 \left(v_{k-1}^T A v_{k-1}^T - \alpha_{k-1} v_{k-1}^T v_{k-1} - \beta_{k-1} v_{k-1}^T v_{k-2} \right) \\ &= \beta_k^- 1 \left(\alpha_{k-1} - \alpha_{k-1} \cdot 1 - 0 \right) \text{ (By IH).} \\ &= 0 \end{aligned}$$

Case 2, show $< v_{k-2}, v_k >= 0$:

$$\begin{aligned} < v_{k-2}, v_k > &= \beta_k^{-1} \left(v_{k-2}^T A v_{k-1}^T - \alpha_{k-1} v_{k-2}^T v_{k-1} - \beta_{k-1} v_{k-2}^T v_{k-2} \right) \\ &= \beta_k^{-1} \left(v_{k-2}^T A v_{k-1} - 0 - \beta_{k-1} \cdot 1 \right) \text{ (By IH).} \end{aligned}$$

Note that:

$$\begin{aligned} \beta_j &= < v_j, \beta v_j > \\ &= v_j^T A v_{j-1} - \alpha_{j-1} v_j^T v_{j-1} - \beta_{j-1} v_j^T v_{j-1} \\ &= v_j^T A v_{j-1} - 0 - 0 \text{ (By IH, assuming } j < k) \\ &= v_j^T A v_{j-1} = v_{j-1}^T A v_j \text{ (A is symmetric)} \end{aligned}$$

Thus,

$$\begin{aligned} < v_{k-2}, v_k > &= \beta_k^{-1} \left(\beta_{k-1} - \beta_{k-1} \right) \\ &= 0 \end{aligned}$$

Case 3, show $< v_j, v_k >= 0, \forall j < k-2$:

$$\begin{aligned} < v_j, v_k > &= \beta_k^{-1} \left(v_j^T A v_{k-1} - \alpha_{k-1} v_j^T v_{k-1} - \beta_{k-1} v_j^T v_{k-2} \right) \\ &= \beta_k^{-1} \left(v_j^T A v_{k-1} - \alpha_{k-1} \cdot 0 - \beta_{k-1} \cdot 0 \right) \text{ (By IH)} \\ &= \beta_k^{-1} \left(v_j^T A v_{k-1} \right) \end{aligned}$$

Need to show that $v_{k-1}^T A v_j = 0, \forall j < k-2$. Expanding the LHS:

$$\begin{aligned} v_{k-1}^T v_j &= v_{k-1}^T A v_{j-1} - \alpha_{i-1} v_{k-1}^T v_{j-1} - \beta_{i-1} v_{k-1}^T v_{j-2} \\ &= v_{k-1}^T A v_{j-1} - 0 - 0 \text{ (By IH)} \\ &= v_{k-1}^T A v_j \\ &= v_j A v_{k-1}^T \text{ (By symmetry)} \\ v_j^T A v_{k-1} &= 0 \end{aligned}$$

Thus:

$$\begin{aligned} < v_j, v_k > &= \beta_k^{-1} \cdot 0 \\ &= 0 \end{aligned}$$

Therefore, by PI, $< v_i, v_j >= \delta_i, \forall i, j$, and the vectors v_j make up an orthonormal set.

Problem 3b: Prove that $v_j = 0$ for some $j \leq n+1$.

Solution 3b: $\{v_j\}$ are an othornormal basis for $\{v, Av, A^2 v \ldots\}$. ????

Problem 3c: Prove that if $v_{j+1} = 0$, thet the eigen values of the tridiagonal matrix T_j are also eigenvalues of A.

Solution 3c: If $V = \{v_1, v_2 \ldots v_j\}$, and $v_{j+1} = 0$, then:

$$AV = VT_j$$

If λ is an eigenvalue of T_j, and x is the associated eigenvector, then:

$$\begin{aligned} T_j x &= \lambda x \\ VT_j x &= \lambda V x \\ AVx &= \lambda V x \end{aligned}$$

Let $w = Vx$:

$$Aw = \lambda w$$

Thus, λ is also an eigenvalue of A and $w = Vx$ is the associated eigenvector.

Problem 4: Give a geometric interpretation of the result of applying the Householder transformation

Solution 4: Rewriting the Householder transformation as:

$$\begin{aligned} T(x) &= x - 2vv^T x \\ &= x - 2v < v, x > \end{aligned}$$

If x is represented in the basis $\{v, u_2 \ldots u_n\}$, where $u_2 \ldots u_n$ are an orthonormal basis of the (n-1) dimensional subspace orthogonal to v, that is, $x = c_1 v + \sum_i = 2^n c_i u_i$:

$$\begin{aligned} T(x) &= c_1 v + \sum_i = 2^n c_i u_i - 2v < v, c_1 v + \sum_i = 2^n c_i u_i > \\ &= c_1 v + \sum_i = 2^n c_i u_i - 2v \cdot (c_1 + 0) \text{ (v is orthogonal to u_i)} \\ &= -c_1 v + \sum_i = 2^n c_i u_i \end{aligned}$$

The result is identical to x except that the component of x in the direction of v has been flipped. Geometrically, this is the reflection of x through the hyperplane whose normal is v, and is spanned by the basis u_i.

Problem 5: For a square matrix A, show how the power method can be generalized to find an interior eigenvalue λ given some estimate σ of that eigenvalue. Describe under which conditions the method is guaranteed to find λ.

Solution 5: Let $\sigma = \lambda - \epsilon$, that is, the true eigenvalue, plus an unknown shift. If v is the eigen vector associated with λ, then:

$$\begin{aligned} Av &= \lambda v \\ Av &= (\sigma + \epsilon) v \\ Av - \sigma v &= \epsilon v \\ (A - \sigma I) v &= \epsilon v \end{aligned}$$

Thus, ϵ is an eigenvalue of $(A-\sigma I)$. Furthermore, since it assumed that λ is the closest eigenvalue to σ, then the desired ϵ is the minimum eigenvalue of $(A-\sigma I)$. The minimum eigenvalue of this matrix can be derived using the Power Method on the inverse of $A-\sigma I$, which of course requires this matrix be invertible for this method to work.

Problem 6: The matrix A's first two columns, a_1, a_2 are such that

$$|a_1^T a_2| \geq ||a_1||_2 ||a_2||_2 (1-\epsilon)$$

Prove that either A is singular or $\kappa(A) \geq \frac{1}{\sqrt{\eta}}$.

Solution 6:

$$\begin{aligned}
\|A\|_2 &= max_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} \\
&= max_{x\neq 0}\sqrt{\left[\frac{\|Ax\|_2}{\|x\|_2}\right]} \\
&= max_{x\neq 0}\sqrt{\left[\frac{<Ax, Ax>}{<x, x>}\right]} \\
&= max_{x\neq 0}\sqrt{\left[\frac{x^T A^T A x}{x^T x}\right]} \\
&= \sqrt{max_{x\neq 0}\frac{x^T A^T A x}{x^T x}} \\
&= \sqrt{max_{x\neq 0}\mathcal{R}_{A^T A}(x)} \\
&= \sqrt{\lambda_{max}(A^T A)}
\end{aligned}$$

$$\begin{aligned}
\left\|A^{-1}\right\|_2 &= max_{x\neq 0}\frac{\left\|A^{-1}x\right\|_2}{\|x\|_2} \\
&= max_{x\neq 0}\sqrt{\left[\frac{\|A^{-1}x\|_2}{\|x\|_2}\right]} \\
&= \sqrt{max_{x\neq 0}\mathcal{R}_{A^{-T}A^{-1}}(x)} \\
&= \sqrt{\lambda_{max}(A^{-T}A^{-1})} \\
&= \sqrt{\frac{1}{\lambda_{min}(AA^T)}}
\end{aligned}$$

Note that AA^T and $A^T A$ have the same eigenvalues:

$$AA^T v = \lambda v$$
$$A^T A(A^T v) = \lambda A^T v$$

Substituting $w = A^T v$:

$$A^T A w = \lambda w$$

Thus:

$$\left\|A^{-1}\right\|_2 = \frac{1}{\sqrt{\lambda_{min}(A^TA)}}$$

The condition number κ of the matrix A is therefore:

$$\begin{aligned}\kappa(A) &= \|A\|_2 \left\|A^{-1}\right\|_2 \\ &= \sqrt{\frac{\lambda_{max}(A^TA)}{\lambda_{min}(A^TA)}}\end{aligned}$$

$$\begin{aligned}A^TA &= \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \dots & \\ - & a_n^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \dots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \dots & a_2^Ta_n \\ . & & & . \\ . & & & . \\ . & & & . \\ a_n^Ta_1 & a_n^Ta_2 & \dots & a_n^Ta_n \end{bmatrix}\end{aligned}$$

In the 2D case:

$$\begin{aligned}\lambda_{max} &= \|a_1\|_2^2 + \|a_2\|_2^2 + \sqrt{\|a_1\|_2^4 + \|a_2\|_2^4 - 2\,\|a_1\|_2^2\,\|a_2\|_2^2 + 4a_1^Ta_2} \\ \lambda_{min} &= \|a_1\|_2^2 + \|a_2\|_2^2 - \sqrt{\|a_1\|_2^4 + \|a_2\|_2^4 - 2\,\|a_1\|_2^2\,\|a_2\|_2^2 + 4a_1^Ta_2}\end{aligned}$$

How to proceed?