Math 330 Linear Algebra Homework 13

12.3 *Least angle property of least squares.* Suppose the $m \times n$ matrix A has linearly independent columns, and b is an m-vector. Let $\hat{x} = A^\dagger b$ denote the least squares approximate solution of $Ax = b$.

(a) Show that for any n-vector x, $(Ax)^T b = (Ax)^T (A\hat{x})$, *i.e.*, the inner product of Ax and b is the same as the inner product of Ax and $A\hat{x}$. *Hint.* Use $(Ax)^T b = x^T (A^T b)$ and $(A^T A)\hat{x} = A^T b$.

(b) Show that when $A\hat{x}$ and b are both nonzero, we have

$$\frac{(A\hat{x})^T b}{\|A\hat{x}\| \|b\|} = \frac{\|A\hat{x}\|}{\|b\|}.$$

The left-hand side is the cosine of the angle between $A\hat{x}$ and b. *Hint.* Apply part (a) with $x = \hat{x}$.

(c) *Least angle property of least squares.* The choice $x = \hat{x}$ minimizes the distance between Ax and b. Show that $x = \hat{x}$ also minimizes the angle between Ax and b. (You can assume that Ax and b are nonzero.) *Remark.* For any positive scalar α, $x = \alpha\hat{x}$ also minimizes the angle between Ax and b.

(a) Compute

$$(Ax)^T(A\hat{x}) = (Ax)^T A A^\dagger b = (Ax)^T A (A^T A)^{-1} A^T b$$

$$= x^T \cancel{A^T A} \cancel{(A^T A)^{-1}} A^T b = x^T A^T b$$

$$= (Ax)^T b$$

(b) Setting $x = \hat{x}$ in the above identity yields

$$\|A\hat{x}\|^2 = (A\hat{x})^T b$$

Now dividing by $\|A\hat{x}\| \|b\|$ on both sides gives

$$\frac{(A\hat{x})^T b}{\|A\hat{x}\| \|b\|} = \frac{\|A\hat{x}\|}{\|b\|}.$$

#12.3 continues...

(c) Since cosine is a decreasing function of angle on the interval $[0, \pi]$ then showing $x = \hat{x}$ is the choice which minimizes the angle between Ax and b is equivalent to showing the choice $x = \hat{x}$ maximizes the cosine of the angle between Ax and b.

The cosine of the angle between Ax and b is given by

$$\frac{(Ax)^T b}{\|Ax\| \, \|b\|}$$

Estimating by Cauchy-Schwarz inequality yields

$$\frac{(Ax)^T b}{\|Ax\| \|b\|} = \frac{(Ax)^T (A\hat{x})}{\|Ax\| \|b\|} \leq \frac{\cancel{\|Ax\|} \|A\hat{x}\|}{\cancel{\|Ax\|} \|b\|}$$

$$= \frac{\|A\hat{x}\|}{\|b\|} = \frac{(A\hat{x})^T b}{\|A\hat{x}\| \|b\|}$$

which is the cosine of the angle between $A\hat{x}$ and b. Therefore we have shown that $x = \hat{x}$ is the choice which minimizes the angle between Ax and b.

Math 330 Linear Algebra Homework 13

12.5 *Approximate right inverse.* Suppose the tall $m \times n$ matrix A has linearly independent columns. It does not have a right inverse, *i.e.*, there is no $n \times m$ matrix X for which $AX = I$. So instead we seek the $n \times m$ matrix X for which the residual matrix $R = AX - I$ has the smallest possible matrix norm. We call this matrix the *least squares approximate right inverse* of A. Show that the least squares right inverse of A is given by $X = A^\dagger$. *Hint.* This is a matrix least squares problem; see page 233.

Since
$$\|R\|^2 = \|AX - I\|^2$$
$$= \|Ax_1 - e_1\|^2 + \cdots + \|Ax_m - e_m\|^2$$

where x_i's are the columns of X and e_i's are the standard basis vectors in $\mathbb{R}^m$, we can minimize R by finding the minimum of $\|Ax_i - e_i\|$ for $i = 1, \ldots, m$ independently.

To minimize $\|Ax_i - e_i\|$ we note that $\hat{x}_i = A^\dagger e_i$.
Therefore the minimizing matrix $\hat{X}$ is given by.

$$\hat{X} = \left[\hat{x}_1 \middle| \hat{x}_2 \middle| \cdots \middle| \hat{x}_m\right] = \left[A^\dagger e_1 \middle| A^\dagger e_2 \middle| \cdots \middle| A^\dagger e_m\right]$$
$$= A^\dagger \left[e_1 \middle| e_2 \middle| \cdots \middle| e_m\right] = A^\dagger I = A^\dagger.$$

Math 330 Linear Algebra Homework 13

12.8 *Least squares and QR factorization.* Suppose A is an $m \times n$ matrix with linearly independent columns and QR factorization $A = QR$, and b is an m-vector. The vector $A\hat{x}$ is the linear combination of the columns of A that is closest to the vector b, *i.e.*, it is the projection of b onto the set of linear combinations of the columns of A.

(a) Show that $A\hat{x} = QQ^Tb$. (The matrix QQ^T is called the *projection matrix*.)

(b) Show that $\|A\hat{x} - b\|^2 = \|b\|^2 - \|Q^Tb\|^2$. (This is the square of the distance between b and the closest linear combination of the columns of A.)

(a) By definition $\hat{x} = A^\dagger b = (A^TA)^{-1}A^Tb$. Plugging in the QR factorization yields that

$$\hat{x} = ((QR)^TQR)^{-1}(QR)^Tb = (R^T\cancel{Q^TQ}R)^{-1}R^TQ^Tb$$

$$= (R^TR)^{-1}R^TQ^Tb$$

and since R is invertible, then

$$\hat{x} = R^{-1}(\cancel{R^T})^{-1}\cancel{R^T}Q^Tb = R^{-1}Q^Tb.$$

It follows that

$$A\hat{x} = AR^{-1}Q^Tb = Q\cancel{R}\cancel{R^{-1}}Q^Tb = QQ^Tb.$$

(b) Calculating yields

$$\|A\hat{x} - b\|^2 = \|QQ^Tb - b\|^2 = (QQ^Tb - b)^T(QQ^Tb - b)$$

$$= (QQ^Tb)^TQQ^Tb - b^TQQ^Tb - (QQ^Tb)^Tb + b^Tb$$

$$= b^TQ\cancel{Q^TQ}Q^Tb - 2b^TQQ^Tb + b^Tb$$

$$= -b^TQQ^Tb + b^Tb = \|b\|^2 - (Q^Tb)^TQ^Tb$$

$$= \|b\|^2 - \|Q^Tb\|^2.$$

Math 330 Linear Algebra Homework 13

12.10 *Numerical check of the least squares approximate solution.* Generate a random 30×10 matrix A and a random 30-vector b. Compute the least squares approximate solution $\hat{x} = A^{\dagger}b$ and the associated residual norm squared $\|A\hat{x} - b\|^2$. (There may be several ways to do this, depending on the software package you use.) Generate three different random 10-vectors d_1, d_2, d_3, and verify that $\|A(\hat{x} + d_i) - b\|^2 > \|A\hat{x} - b\|^2$ holds. (This shows that $x = \hat{x}$ has a smaller associated residual than the choices $x = \hat{x} + d_i$, $i = 1, 2, 3$.)

Solution is in Julia on the next page...

hw13p1210

December 3, 2020

Homework 13 Question 12.10

[18]:
```
using LinearAlgebra
```

[19]:
```
A=rand(30,10);
b=rand(30);
```

[20]:
```
Adagger=inv(A'*A)*A';
```

[21]:
```
xhat=Adagger*b;
```

[22]:
```
norm(A*xhat-b)
```

[22]: 1.494082025485521

[23]:
```
d1=rand(10);
d2=rand(10);
d3=rand(10);
```

[24]:
```
norm(A*(xhat+d1)-b)
```

[24]: 15.498057939880987

[25]:
```
norm(A*(xhat+d2)-b)
```

[25]: 14.142275011244568

[26]:
```
norm(A*(xhat+d3)-b)
```

[26]: 14.427812725660111

Note in each of these cases that

$$\|A(\hat{x} - d_i) - b\| > \|A\hat{x} - b\|.$$

We try again with random vectors d_i with smaller norms as the random vectors chosen above are relatively large compared to the least squares solution $\tilde{x}$.

[27]:
```
d1=1e-6*rand(10);
d2=1e-6*rand(10);
d3=1e-6*rand(10);
```

We then compute

$$\|A(\hat{x} - d_i) - b\| - \|A\hat{x} - b\|$$

for each value of i and check that this difference is positive.

[28]:
```
norm(A*(xhat+d1)-b)-norm(A*xhat-b)
```

[28]: 3.779554447191913e-11

[29]:
```
norm(A*(xhat+d2)-b)-norm(A*xhat-b)
```

[29]: 1.3354650718611083e-10

[30]:
```
norm(A*(xhat+d3)-b)-norm(A*xhat-b)
```

[30]: 6.07782713046845e-11

Though small, each of the differences are positive. This constitutes a numerical verification of the least squares approximate solution.

[ ]:
```

```