Math221 Homework # 1 Solutions

Instructor: James W. Demmel

September 19, 2009

1.1.

Let A be an orthogonal matrix. Show that $\det(A) = \pm 1$. Show that if B is also orthogonal and $\det(A) = -\det(B)$, then $A + B$ is singular.

Solution: $(\det A)^2 = \det A \det A = \det A \det A^T = \det AA^T = \det I = 1$

First proof that $A + B$ is singular: $A + B$ is singular iff $A^T(A + B) = I + A^T B$ is. $A^T B$ is real and orthogonal so if λ is an eigenvalue and x is its eigenvector, $\|x\|_2 = \|(A^T B)x\|_2 = \|\lambda x\|_2 = |\lambda| \cdot \|x\|_2$, and so $|\lambda| = 1$. Thus the eigenvalues are 1, -1, or complex conjugate pairs λ and $\bar{\lambda}$ of absolute value 1. The product of all the complex eigenvalues is 1 (since $\lambda\bar{\lambda} = |\lambda|^2 = 1$). Since the product of all the eigenvalues is equal to $\det A^T B = \det A \det B = -1$ then at least one of the eigenvalues of $A^T B$ must be -1. Let the corresponding vector be x. Then $(I + A^T B)x = x - x = 0$, so $I + A^T B$ is singular and so is $A + B$.

Second proof: $\det(A + B) = -\det(A^T)\det(A + B)\det(B^T) = -\det(A^T AB^T + A^T BB^T) = -\det(A^T + B^T) = -\det(A + B)$, so $\det(A + B) = 0$.

1.2.

The rank of a matrix is the dimension of the space spanned by its columns. Show that if A has rank one if and only if $A = ab^T$ for some column vectors a and b.

Solution: If $A = (A_{ij})$ has rank 1, pick a nonzero row (say) row i. Then all other rows are multiples of row i. Let $[b_1, ..., b_n]$ be row i of A. In other words row k is $[A_{k1}, A_{k2}, ..., A_{kn}]$ and must satisfy $A_{kj} = a_k \cdot b_j$ for some scalar a_k. This may also be written $A = ab^T$, where the entries of a are a_k and the entries of b are b_j. Conversely, if $A = ab^T$, then the space spanned by A's columns is the one-dimensional space of all multiples of a.

1.3.

Show that if a matrix is orthogonal and triangular then it is diagonal. What are its diagonal entries?
Solution 1: If $A = (a_{ij})$ is orthogonal and (say upper) triangular ($a_{ij} = 0, i > j$) then $AA^T = I$. The last equality yields ($A(:, n)A^T = e_n$): $a_{nn}a_{1n} = 0, a_{nn}a_{2n} = 0, ..., a_{nn}a_{n-1,n} = 0, a_{nn}^2 = 1$ so a_{nn} is equal to either 1 or -1 and $a_{1n} = a_{2n} = ... = a_{n-1,n} = 0$. Proceeding by induction we get that A is diagonal and the diagonal entries are +1 or -1.

Solution 2: Assume A is upper triangular. Then A^{-1} is also upper triangular. But $A^{-1} = A^T$ is lower triangular. So A must be diagonal. Since A is orthogonal, $I = A^T A$ and so the squares of all the diagonal entries of A must be 1, and the entries themselves are ± 1.

1.4.

A matrix is strictly upper if it is upper triangular with zero diagonal elements. Show that if A is strictly upper triangular and $n \times n$, then $A^n = 0$.

Solution: First proof: Use induction to show that if $A^k = (a_{ij}^{(k)})$ then $a_{ij}^{(k)} = 0$ if $i > j - k$. Thus $a_{ij}^{(n)} = 0$ for $i > j - n$, i.e. for all $i, j = 1, 2, ..., n$. So $A^n = 0$.

Second proof: All eigenvalues of A are 0 so its characteristic polynomial is $\det(A - xI) = x^n$. By the Cayley-Hamilton theorem, A satisfies its own characteristic polynomial $A^n = 0$.

1.5

Suppose that $\| \cdot \|$ is a vector norm, and that C is m-by-n with rank n. Show that $\|x\|_C \equiv \|Cx\|$ is also a vector norm.

We need to check all 3 parts of Definition 1.3 of a vector norm:

1. $\|x\|_C = \|Cx\| \geq 0$ since $\|y\| \geq 0$ for all y. Similarly, $\|Cx\| = 0$ if and only if $Cx = 0$, but since C has full column rank, $Cx = 0$ if and only if $x = 0$.
2. $\|\alpha x\|_C = \|\alpha Cx\| = |\alpha| \cdot \|Cx\| = |\alpha| \cdot \|x\|_C$.
3. $\|x + y\|_C = \|C(x + y)\| = \|Cx + Cy\| \leq \|Cx\| + \|Cy\| = \|x\|_C + \|y\|_C$.

1.10.

We compute $x^T y$ as follows:

```
$s = x_1 \cdot y_1$
for $i = 2 : d$
    $s = s + x_i \cdot y_i$
end
```

Incorporating roundoff errors, we get

```
$\hat{s} = x_1 \cdot y_1(1 + \delta_1)$
for $i = 2 : d$
    $\hat{s} = (\hat{s} + x_i \cdot y_i(1 + \delta_i))(1 + \delta_i')$
end
```

(Note: the book abbrebiated this by writing $fl(x^T y)$ to mean the final value of $\hat{s}$.) Therefore we have that

$$\begin{aligned} fl(x^T y) &= (\ldots((x_1 y_1(1 + \delta_1) + x_2 y_2(1 + \delta_2))(1 + \delta_2') + x_3 y_3(1 + \delta_3))(1 + \delta_3') \ldots \\ &= \sum_{i=1}^{d} x_i y_i \left[(1 + \delta_i) \prod_{j=i}^{d} (1 + \delta_j') \right], \end{aligned}$$

where we define $\delta_1' = 0$. All δ's are bounded by ε in absolute value. Using the fact that $|\delta_i| \leq \varepsilon$ implies that $(1 + \delta_1)...(1 + \delta_s) = (1 + s\delta)$ for some $|\delta| < \varepsilon$ (here and elsewhere we ignore terms bounded by $O(\varepsilon^2)$) we may write

$$fl(x^T y) = \sum_{i=1}^{d} x_i y_i (1 + d\delta_i),$$

where $|\delta_i| < \varepsilon$. This formula is true no matter in what order the $x_i \cdot y_i$ are summed.

For the second part of the problem we have:

$$\begin{aligned} |fl(AB) - AB|_{ij} &= \left| \sum_{k=1}^{n} a_{ik} b_{kj} (1 + n\delta_{ijk}) - \sum_{k=1}^{n} a_{ik} b_{kj} \right| \quad \text{where } |\delta_{ijk}| \leq \varepsilon \\ &= \left| \sum_{k=1}^{n} a_{ik} b_{kj} n \delta_{ijk} \right| \leq \sum_{k=1}^{n} |a_{ik}| \cdot |b_{kj}| \cdot n |\delta_{ijk}| \leq n\varepsilon (|A| \cdot |B|)_{ij}. \end{aligned}$$

1.11.

The precise algorithm is as follows:

```
for $i = 1 : n$
    $x_i = b_i$
    for $k = 1 : i - 1$
        $x_i = x_i - l_{ik} \cdot x_k$
    end
    $x_i = x_i / l_{ii}$
end
```

Incorporating roundoff errors we get

```
for $i = 1 : n$
    $\hat{x}_i = b_i$
    for $k = 1 : i - 1$
        $\hat{x}_i = (\hat{x}_i - l_{ik} \cdot \hat{x}_k (1 + \delta_{ik}))(1 + \delta'_{ik})$
    end
    $\hat{x}_i = (\hat{x}_i / l_{ii})(1 + \delta'_{ii})$
end
```

Altogether, we get that

$$\hat{x}_i = \frac{1}{l_{ii}} \left(b_i \prod_{j=1}^{i-1} (1 + \delta'_{ij}) - \sum_{k=1}^{i-1} l_{ik} \hat{x}_k (1 + \delta_{ik}) \prod_{j=k}^{i-1} (1 + \delta'_{ij}) \right) (1 + \delta'_{ii}).$$

Multiply both sides by $l_{ii} / \prod_{j=1}^{i} (1 + \delta'_{ij})$ and use the fact that $(1 + \delta)^{-1} = (1 - \delta)$ (again ignoring $O(\varepsilon^2)$ terms), to get

$$\sum_{j=1}^{i} l_{ij} (1 + j\delta''_{ij}) \hat{x}_j = b_i$$

where $|\delta''_{ij}| \leq \varepsilon$. That is, the computed solution $\hat{x}$ satisfies a slightly perturbed system

$$(L + \delta L)\hat{x} = b$$

with $|\delta L| < n\varepsilon |L|$, where the inequality is interpresteded componentewise. This bound on $|\delta L|$ is true no matter what order of summation is used in the innermost loop of the algorithm.