Matrix Theory HWs

Liam Flaherty

January 11, 2024

Contents

1 Scalar, Vector, And Matrix Multiplication **3**

2 Floating Point Arithmetic **7**

3 Matrix And Vector Norms **10**

4 Singular Value Decomposition **14**

5 Projections And Least Squares **17**

1 Scalar, Vector, And Matrix Multiplication

Problem 1.1) Suppose $\boldsymbol{Ax} = \boldsymbol{Ay} = \boldsymbol{0}$ for $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n}$. Combine the two statements into a single equation $\boldsymbol{AB} = \boldsymbol{C}$ that is equivalent to the above. Determine $\boldsymbol{B}$ and $\boldsymbol{C}$, and specify their dimensions.

Consider $B = \begin{bmatrix} x & y \end{bmatrix} \in \mathbb{R}^{n\times 2}$ (and thus $C = 0 \in \mathbb{R}^{m\times 2}$).

First assume $Ax = Ay = 0$. Right-multiplying A by B gives a linear combination of the columns; $AB = \begin{bmatrix} AB_{\bullet,1} & AB_{\bullet,2} \end{bmatrix} = \begin{bmatrix} Ax & Ay \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} \in \mathbb{R}^{m\times 2}$; $Ax = Ay = 0 \implies AB = C$.

Next assume $AB = C = 0 \in \mathbb{R}^{m\times 2}$. Since a matrix is zero only when all its entries are zero, and we showed above that $AB = \begin{bmatrix} Ax & Ay \end{bmatrix}$, we must have $Ax = 0$ and $Ay = 0$. This proves the equivalence.

Problem 1.2) Determine the column vectors $\boldsymbol{u}$ and $\boldsymbol{w}$ such that $\begin{bmatrix} \mathbf{1} & \mathbf{-2} & \mathbf{-2} \\ \mathbf{-2} & \mathbf{2} & \mathbf{4} \end{bmatrix} = \boldsymbol{uw^T}$ and $\boldsymbol{u}$ is equal to the second column of the matrix.

We are given $u= \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and want to find w such that $\begin{bmatrix} -1 \\ 2 \end{bmatrix} w^T = \begin{bmatrix} 1 & -2 & -2 \\ -2 & 2 & 4 \end{bmatrix} = A$. One way to look at this is to find the scalar multiples of u that form the columns of A. The first column of A is $A_{\bullet,1} = -u$, $A_{\bullet,2} = u$, and $A_{\bullet,3} = 2u$. So $w^T = \begin{bmatrix} -1 & 1 & 2 \end{bmatrix}$ and we have $u = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

Problem 1.3) For $\boldsymbol{x} \in \mathbb{R}^{n}$ and $\boldsymbol{y} \in \mathbb{R}^{1\times n}$, show how to compute $(\boldsymbol{xy})^3\boldsymbol{x}$ using only inner products and scalar multiplications.

Write $(xy)^3x = (xy)(xy)(xy)x = x(yx)(yx)(yx)x = x(yx)^3$ by the associativity of matrix multiplication. Since $y \in \mathbb{R}^{1\times n}$ and $x \in \mathbb{R}^{n\times 1}$, $yx = \sum_{i=1}^{n} y_{i,1}x_{1,i} = r \in \mathbb{R}$; it is an inner product. Using scalar multiplication, we can compute $(yx)^3 = r^3 = \left(\sum_{i=1}^{n} y_{i,1}x_{1,i}\right)^3$. Then, since scalar quantities commute with matrices, we can write $xr^3 = r^3x = (yx)^3x$ and perform element-wise scalar multiplication on the elements of x with the quantity $(yx)^3$ to get our desired computation.

Problem 1.4) 4) Let $\boldsymbol{x_1, \ldots, x_n, a} \in \mathbb{R}^{\boldsymbol{d}}$ be vectors, and define the matrix $\boldsymbol{C} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{(x_j - a)(x_j - a)^T}$. Derive the expression $\boldsymbol{C = (X - a\mathbb{1}^T)(X - a\mathbb{1}^T)^T}$ where $\mathbb{1}$ is the vector of all ones, and $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{d \times n}}$.

Consider $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \in \mathbb{R}^{d\times n}$. Since $X \in \mathbb{R}^{d\times n}$, we must have $a\mathbb{1}^T \in \mathbb{R}^{d\times n}$ as well, and then since $a \in \mathbb{R}^{d\times 1}$, we must have $\mathbb{1}^T \in \mathbb{R}^{1\times n}$.

Let $(x_j)_i$ denote the i^{th} element of the column vector x_j. Similarly let a_i denote the i^{th} element of the column vector a.

For notational ease, label $(x_j - a)(x_j - a)^T = \begin{bmatrix} (x_j)_1 - a_1 \\ (x_j)_2 - a_2 \\ \vdots \\ (x_j)_d - a_d \end{bmatrix} \begin{bmatrix} (x_j)_1 - a_1 \\ (x_j)_2 - a_2 \\ \vdots \\ (x_j)_d - a_d \end{bmatrix}^T = v_j v_j^T = M_j$.

Now observe:

$$
\begin{aligned}
C &= (X - a\mathbb{1}^T)(X - a\mathbb{1}^T)^T \\
&= \left(\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & \vdots & \downarrow \end{bmatrix} - \begin{bmatrix} a & a & \cdots & a \\ \downarrow & \downarrow & \vdots & \downarrow \end{bmatrix} \right) \left(\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & \vdots & \downarrow \end{bmatrix} - \begin{bmatrix} a & a & \cdots & a \\ \downarrow & \downarrow & \vdots & \downarrow \end{bmatrix} \right)^T \\
&= \left(\underbrace{\begin{bmatrix} (x_1 - a) & (x_2 - a) & \cdots & (x_n - a) \\ \downarrow & \downarrow & \vdots & \downarrow \end{bmatrix}}_{\in \mathbb{R}^{d\times n}} \right) \left(\underbrace{\begin{bmatrix} (x_1 - a) & \rightarrow \\ (x_2 - a) & \rightarrow \\ \cdots & \rightarrow \\ (x_n - a) & \rightarrow \end{bmatrix}}_{\in \mathbb{R}^{n\times d}} \right) \\
&= \begin{bmatrix} \sum_{j=1}^n ((x_j)_1 - a_1)((x_j)_1 - a_1) & \sum_{j=1}^n ((x_j)_1 - a_1)((x_j)_2 - a_2) & \cdots & \sum_{j=1}^n ((x_j)_1 - a_1)((x_j)_d - a_d) \\ \sum_{j=1}^n ((x_j)_2 - a_2)((x_j)_1 - a_1) & \sum_{j=1}^n ((x_j)_2 - a_2)((x_j)_2 - a_2) & \cdots & \sum_{j=1}^n ((x_j)_2 - a_2)((x_j)_d - a_d) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n ((x_j)_d - a_d)((x_j)_1 - a_1) & \sum_{j=1}^n ((x_j)_d - a_d)((x_j)_2 - a_2) & \cdots & \sum_{j=1}^n ((x_j)_d - a_d)((x_j)_d - a_d) \end{bmatrix} \\
&= \sum_{j=1}^n \begin{bmatrix} ((x_j)_1 - a_1)((x_j)_1 - a_1) & ((x_j)_1 - a_1)((x_j)_2 - a_2) & \cdots & ((x_j)_1 - a_1)((x_j)_d - a_d) \\ ((x_j)_2 - a_2)((x_j)_1 - a_1) & ((x_j)_2 - a_2)((x_j)_2 - a_2) & \cdots & ((x_j)_2 - a_2)((x_j)_d - a_d) \\ \vdots & \vdots & \ddots & \vdots \\ ((x_j)_d - a_d)((x_j)_1 - a_1) & ((x_j)_d - a_d)((x_j)_2 - a_2) & \cdots & ((x_j)_d - a_d)((x_j)_d - a_d) \end{bmatrix} \\
&= \sum_{j=1}^n M_j = \sum_{j=1}^n v_j v_j^T = \sum_{j=1}^n (x_j - a)(x_j - a)^T
\end{aligned}
$$

Problem 1.5) For $\boldsymbol{x_j} \in \mathbb{R}^n$, $1 \leq j \leq 4$, determine the matrix $\boldsymbol{Z}$ that produces the transformation $\boldsymbol{Z} \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \boldsymbol{x_3} \\ \boldsymbol{x_4} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_3} \\ \boldsymbol{x_4} \end{bmatrix}$ and represent $\boldsymbol{Z}$ as a sum of outer products.

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$$

Problem 1.6) Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ be a matrix with $\boldsymbol{A} + \frac{1}{3}\boldsymbol{A}^3 + \frac{1}{5}\boldsymbol{A}^5 = \boldsymbol{I}$. Is $\boldsymbol{A}$ singular or nonsingular?

A is nonsingular. See that $A(I + \frac{1}{3}A^2 + \frac{1}{5}A^4) = I$ by the distributive laws and so (by the uniqueness of inverses), $A^{-1} = (I + \frac{1}{3}A^2 + \frac{1}{5}A^4)$, which is well defined because A is square.

Problem 1.7) For the matrix $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{v}(\boldsymbol{w}^T\boldsymbol{v})^{-1}\boldsymbol{w}^T \in \mathbb{R}^{n \times n}$ where $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ with $\boldsymbol{w}^T\boldsymbol{v} \neq 0$, present a derivation for $\boldsymbol{B}^2$.

We have:

$$\begin{aligned}
B^2 &= \left(I - v(w^Tv)^{-1}w^T\right)\left(I - v(w^Tv)^{-1}w^T\right) && \text{Definition} \\
&= I - 2v(w^Tv)^{-1}w^T + v(w^Tv)^{-1}w^Tv(w^Tv)^{-1}w^T && \text{Expanding} \\
&= I - 2v(w^Tv)^{-1}w^T + v(w^Tv)^{-1}(w^Tv)(w^Tv)^{-1}w^T && \text{Grouping} \\
&= I - 2v(w^Tv)^{-1}w^T + v(w^Tv)^{-1}w^T && \text{Properties of inverses} \\
&= I - v(w^Tv)^{-1}w^T && \text{Simplifying} \\
&= B && B \text{ is idempotent}
\end{aligned}$$

Problem 1.8) Show that an upper triangular orthogonal matrix must be diagonal.

Let A be both upper triangular and orthogonal. As A is upper triangular, A^T is lower triangular, i.e. $A^T_{i,j} = 0$ for all $i < j$. On the other hand, A^{-1} is also upper triangular, i.e. $A^{-1}_{i,j} = 0$ for all $i > j$. Combined with the fact that A is orthogonal, i.e. $A^{-1} = A^T$, we must have $A^{-1}_{i,j} = A^T_{i,j} = 0$ whenever $i \neq j$. This is precisely the definition of a diagonal matrix.

Problem 1.9) Let $\boldsymbol{X} = \underbrace{[\boldsymbol{b} \cdots \boldsymbol{b}]}_{\boldsymbol{n} \text{ times}} \in \mathbb{R}^{m \times n}$ where $\boldsymbol{b} \in \mathbb{R}^m$. Present a derivation to determine a simple expression for $\boldsymbol{X}\boldsymbol{X}^T$.

Call b_i the i^{th} element of b. Then:

$$XX^T = \begin{bmatrix} b_1 & \cdots & b_1 \\ \vdots & \ddots & \vdots \\ b_m & \cdots & b_m \end{bmatrix} \begin{bmatrix} b_1 & \cdots & b_m \\ \vdots & \ddots & \vdots \\ b_1 & \cdots & b_m \end{bmatrix} = \begin{bmatrix} nb_1^2 & nb_1b_2 & \cdots & nb_1b_m \\ nb_2b_1 & nb_2^2 & \cdots & nb_2b_m \\ \vdots & \vdots & \ddots & \vdots \\ nb_mb_1 & nb_mb_2 & \cdots & nb_m^2 \end{bmatrix} = nbb^T$$

Problem 1.10) For vectors $x, y \in \mathbb{R}^n$ show that $x^T y = \frac{1}{4}(\|x+y\|_2^2 - \|x-y\|_2^2)$.

First, recall that for any vector $a \in \mathbb{R}^n$, $\|a\|_2^2 = a^T a$. Next, recall that for any vectors $a, b \in \mathbb{R}^n$, $a^T b = b^T a$.

Proceeding to the problem at hand, we can write:

$$\begin{aligned}
\|x+y\|_2^2 &= (x+y)^T(x+y) && \text{First note above} \\
&= (x^T + y^T)(x+y) && \text{Transpose rules} \\
&= x^T x + x^T y + y^T x + y^T y && \text{Distributing} \\
&= \|x\|_2^2 + x^T y + y^T x + \|y\|_2^2 && \text{First note above} \\
&= \|x\|_2^2 + \|y\|_2^2 + 2x^T y && \text{Second note above}
\end{aligned}$$

And similarly:

$$\begin{aligned}
\|x-y\|_2^2 &= (x-y)^T(x-y) && \text{First note above} \\
&= (x^T - y^T)(x-y) && \text{Transpose rules} \\
&= x^T x - x^T y - y^T x + y^T y && \text{Distributing} \\
&= \|x\|_2^2 - x^T y - y^T x + \|y\|_2^2 && \text{First note above} \\
&= \|x\|_2^2 + \|y\|_2^2 - 2x^T y && \text{Second note above}
\end{aligned}$$

Then $(\|x+y\|_2^2 - \|x-y\|_2^2) = (\|x\|_2^2 + \|y\|_2^2 + 2x^T y) - (\|x\|_2^2 + \|y\|_2^2 - 2x^T y) = 4x^T y$. Dividing through by 4, we reach our result.

2 Floating Point Arithmetic

Problem 2.1) The expressions 2^{-1024} and $\frac{1}{2^{1024}}$ are equal in exact arithmetic. Explain precisely what happens when they are evaluated in IEEE double precision floating point arithmetic (in MATLAB), and why.

We first note that the normalized double-precision floating point of a number is given by $(-1)^s(1.f)_2 2^{e-1023}$ where s is allocated 1 bit, f is allocated 52 bits, and e is allocated 11 bits. Since e is allocated 11 bits, (and since cases of all "1"'s or all "0"'s are reserved for special values), the total exponent after including the bias is minimally $1-1023 = -1022$ and maximally $(2^{11}-1-1)-1023 = 1023$. Thus, $\texttt{realmin} = 2^{-1022}$ and $\texttt{realmax} < 2 \cdot 2^{1023} = 2^{1024}$ (the smallest and largest normalized representation of numbers, respectively).

In the case of $\frac{1}{2^{1024}}$, MATLAB first computes 2^{1024}, and then takes the reciprocal of the result. But since $2^{1024} > \texttt{realmax}$ by the above explanation, 2^{1024} results in overflow and evaluates as `Inf`. In contrast to `NaN`, operations with `Inf` can still result in real numbers in MATLAB. We see that here, as $\frac{1}{2^{1024}}$ is interpreted as $1/\texttt{Inf} = 0$ exactly.

In the case of 2^{-1024}, since $\left[\texttt{eps}(0) = (0.\underbrace{0\ldots0}_{51\text{ times}}1)_2 2^{-1022}\right] < 2^{-1024} < \texttt{realmin}$, the evaluation of 2^{-1024} results in gradual underflow. It is therefore represented as a denormalized number (i.e., without the leading "1" bit preceding f), and is not exactly zero.

```
>> 2^1024

ans =

   Inf

>> 1/(2^1024)

ans =

     0

>> 2^(-1024)

ans =

  5.5627e-309
```

Figure 2.1: MATLAB Output

Problem 2.2) Determine the exact values of `x` and `xp1`, in terms of powers of two, at the termination of the MATLAB algorithm below. Explain the purpose of the algorithm.

```
x = 1;
xp1= x + 1;
while xp1 > 1
x=x/2;
xp1=x+1;
end
x=2x
```

At termination, `x` is machine epsilon (it is $\texttt{x} \approx \texttt{eps} = 2^{-52}$), and `xp1` is exactly $1 + \frac{\texttt{eps}}{2}$ (but 1 in floating point). Apparently, the purpose of this algorithm is to determine the order of values of y such that the computer cannot distinguish 1 from $1 + y$.

Since multiplying normalized floating point numbers by powers of two results in exact values, the while loop stores a repeatedly smaller exact value. The loop simply divides x by 2 after each iteration, so the fractional part of the number does not change and there are no roundoff errors. Before the `while` loop begins, $x = (-1)^0(1.\underbrace{0\ldots0}_{\text{52 times}})_2 = 2^{1023-1023} = 2^0$, after the first loop, $x = 2^{-1}$, after the second loop, $x = 2^{-2}$, etc.

This process continues for an long as the computer evaluates $\texttt{xp1} \equiv 1+2^{-n}$ (for $n \in \mathbb{N}$) as strictly greater than 1. By the definition of machine epsilon, `eps` is the next largest floating point number after 1, and so the computer cannot store distinct floating point numbers between 1 and $1 + \texttt{eps}$. As such, after the 53^{rd} iteration of the `while` loop, the process is exited; the computer evaluates $1 + 2^{-53}$ as equal to 1. After multiplying `x` by 2 at the conclusion of the loop, we reach the aforementioned result.

Problem 2.3) Determine the quantities f and e in the normalized representation of the number 33, then determine the distance of 33 to the next larger floating point number.

We know we can write $33 = (-1)^s(1.f)_2 \cdot 2^{e-1023}$ where the sign s is allocated 1 bit, the fraction f is 52 binary bits, and the (biased) exponent e is 11 binary bits. Since 33 is positive, $s = 0$.

To get the approximate magnitude of the number, we note that $2^5 = 32 < 33$, so our unbiased exponent should be 5, and thus $e = 1028$ in decimal (and so $1\underbrace{0\cdots0}_{7\text{ times}}100$ in binary).

We then scale the magnitude with the significant $(1.f)_2$. We need $(1.f)_2 = \frac{33}{32}$ and so $f = \frac{1}{32} = 2^{-5}$ (which is $00001\underbrace{0\cdots0}_{47\text{ times}}$ in binary)

So $33 = (-1)^0(1+2^{-5})2^5$. The next largest floating point number is an increment of `eps` $= 2^{-52}$ in the fraction; we have $(1.f)_2 = 00001\underbrace{0\cdots0}_{46\text{ times}}1$ in the new number and so it is $(-1)^0(1+2^{-5}+2^{-52})2^5 = 33+2^{-47}$; our increment is 2^{-47}.

3 Matrix And Vector Norms

Problem 3.1) For $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$, show that $\|\boldsymbol{x}\|_{\mathbf{2}} \leq \sqrt{\boldsymbol{n}}\|\boldsymbol{x}\|_{\infty}\|\boldsymbol{x}\|_{\mathbf{1}}$.

Let $x= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ be given. For any of the x_j, $|x_j|^2 \leq |x_j||x_j| \leq \left(\max\limits_{1\leq i\leq n} |x_i|\right) |x_j|$. Then:

$$
\begin{aligned}
\sum_{j=1}^{n} |x_j|^2 &\leq \sum_{j=1}^{n} \left(\max_{1\leq i\leq n} |x_i|\right) |x_j| && \text{Above fact} \\
\sum_{j=1}^{n} |x_j|^2 &\leq \left(\max_{1\leq i\leq n} |x_i|\right) \sum_{j=1}^{n} |x_j| && \text{Distributive Law} \\
\sum_{j=1}^{n} |x_j|^2 &\leq \|x\|_\infty \sum_{j=1}^{n} |x_j| && \text{Definition of } \|x\|_\infty \\
\sum_{j=1}^{n} |x_j|^2 &\leq \|x\|_\infty \|x\|_1 && \text{Definition of } \|x\|_1 \\
\sqrt{\sum_{j=1}^{n} |x_j|^2} &\leq \sqrt{\|x\|_\infty \|x\|_1} && \text{Both sides positive} \\
\|x\|_2 &\leq \sqrt{\|x\|_\infty \|x\|_1} && \text{Definition of } \|x\|_2
\end{aligned}
$$

Problem 3.2) Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be nonsingular, with pivoted $\boldsymbol{LU}$ decomposition $\boldsymbol{PA} = \boldsymbol{LU}$ where $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is a permutation matrix, $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ is upper triangular, and $\boldsymbol{L}$ is unit lower triangular with all elements $|\boldsymbol{L_{i,j}}| \leq 1$. Show that the infinity norm of $\boldsymbol{U^{-1}}$ can be bounded in terms of the infinity norm of $\boldsymbol{A^{-1}}$, $\|\boldsymbol{U^{-1}}\|_\infty \leq \boldsymbol{n}\|\boldsymbol{A^{-1}}\|_\infty$.

All permutation matrices are invertible (all rows are linearly independent since, by the definition of invertibility, each row has exactly one non-zero entry, and no other entries in the same column are non-zero). We are given that A is nonsingular. So PA can be brought to row echelon form U (this is a necessary condition of invertibility) by a series of elementary matrices multiplying to L^{-1}, and we know that all four matrices are invertible.

With this insight, we can write $U^{-1} = A^{-1}P^{-1}L$ by the rules for matrix inverses. Observe:

$$
\begin{aligned}
\|U^{-1}\|_\infty &= \|A^{-1}P^{-1}L\|_\infty && \text{Applying norm to above} \\
&\leq \|A^{-1}\|_\infty\|P^{-1}L\|_\infty && \text{Submultiplicative Property and commutativity} \\
&= \|A^{-1}\|_\infty\|P^{-1}\|_\infty\|L\|_\infty && P \text{ is a permutation matrix, so orthogonal} \\
&= \|A^{-1}\|_\infty\|L\|_\infty && \text{Row sums of permutation matrix are all 1} \\
&\leq n\|A^{-1}\|_\infty && \text{We've assumed } |L_{i,j}| \leq 1\text{, so the max row sum of } L \text{ is } n
\end{aligned}
$$

Problem 3.3) Let $A = [a_1, \dots, a_n] \in \mathbb{R}^{m \times n}$ have columns $a_j \in \mathbb{R}^m$. Show that the two-norm of A can bounded in terms of the largest column two-norm, $\|A\|_2 \leq \sqrt{n} \max_{1 \leq j \leq n} \|a_j\|_2$. Your proof should use the definition of the two-norm, and be presented in terms of columns rather than individual matrix elements.

By the definition of matrix norm, $\|A\|_2 = \max_{\|y\|_2 = 1} \|Ay\|_2$. Choose $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$ and $\|Ax\|_2 = \|A\|_2$. Now observe:

$$\begin{aligned}
\|A\|_2 &= \|Ax\|_2 && \text{Definition} \\
&= \left\| \sum_{j=1}^{n} x_j a_j \right\|_2 && \text{Column view of matrix multiplication} \\
&\leq \sum_{j=1}^{n} \|x_j a_j\|_2 && \text{Triangle Inequality} \\
&\leq \sum_{j=1}^{n} |x_j| \cdot \|a_j\|_2 && \text{Homogeneity} \\
&\leq \left(\sum_{j=1}^{n} |x_j| \right) \max_{1 \leq j \leq n} \|a_j\|_2 && \text{Distributive Law after bounding by maximum} \\
&\leq \|x\|_2 \sqrt{n} \cdot \max_{1 \leq j \leq n} \|a_j\|_2 && \text{Cauchy-Schwarz Inequality, } \|x\|_1 \leq \sqrt{n}\|x\|_2 \\
&\leq \sqrt{n} \cdot \max_{1 \leq j \leq n} \|a_j\|_2 && \text{How } x \text{ was defined}
\end{aligned}$$

The second to last inequality comes from the Cauchy-Schwarz inequality, that $|a^T b| \leq \|a\|_2 \|b\|_2$ for all $a, b \in \mathbb{R}^n$, and then taking a to be our choice of x and b to be the column vector of all 1's. So $\|x^T b\|_1 = \|x\|_1 \leq \|x\|_2 \sqrt{\sum_{j=1}^{n} b_j^2} = \|x\|_2 \sqrt{\sum_{j=1}^{n} 1} = \|x\|_2 \cdot \sqrt{n}$ and we've proved our desired result.

Problem 3.4) Let the nonsingular matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be partitioned as $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$ where $\boldsymbol{A}_{11} \in \mathbb{R}^{k\times k}$ for some $1 \leq k \leq n$ and $\boldsymbol{A}_{11}$ is nonsingular. Very that $\boldsymbol{A}$ has a block $\boldsymbol{LU}$ factorization, $\boldsymbol{L} = \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{I}_{n-k} \end{bmatrix}$ and $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{U}_{22} \end{bmatrix}$ where $\boldsymbol{U}$ is block upper triangular, by expressing $\boldsymbol{L}_{21}$ and $\boldsymbol{U}_{22}$ in terms of the blocks of $\boldsymbol{A}$.

First note the dimensions of the blocks. Since $L \in \mathbb{R}^{n\times n}$ and since I_k has k rows, L_{21} must have $n-k$ rows. Further, since I_{n-k} has $n-k$ columns, L_{21} must have k columns. We can apply this same logic for the two other matrices to see $A_{12} \in \mathbb{R}^{k\times(n-k)}$, $U_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$, and $A_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$. This is a useful check to verify the below computations are valid.

With this check out the way, we now expand the factorization:

$$
\begin{aligned}
A &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \\
&= LU \\
&= \begin{bmatrix} I_k & 0 \\ L_{21} & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ 0 & U_{22} \end{bmatrix} \\
&= \begin{bmatrix} I_k(A_{11}) + 0(0) & I_k(A_{12}) + 0(U_{22}) \\ L_{21}(A_{11}) + I_{n-k}(0) & L_{21}(A_{12}) + I_{n-k}(U_{22}) \end{bmatrix} \\
&= \begin{bmatrix} A_{11} & A_{12} \\ L_{21}A_{11} & L_{21}(A_{12}) + U_{22} \end{bmatrix}
\end{aligned}
$$

This shows that we need $A_{21} = L_{21}A_{11}$ and subsequently $A_{22} = L_{21}(A_{12}) + U_{22}$.

We are given that A_{11} is invertible, so have $L_{21} = A_{21}A_{11}^{-1}$, which is well-defined. Plugging this value in to the formula for U_{22}, we see $U_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ and have reached our result.

4 Singular Value Decomposition

Problem 4.1) Let $A \in \mathbb{R}^{m \times n}$ with $m > n$ and $\mathrm{Dim}(C(A)) = n$. Use a singlular value decomposition (SVD) of A to determine a SVD of $A^T A$.

Consider the SVD $A = U\Sigma V$ where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, and $\Sigma = \begin{bmatrix}\Sigma \\ 0\end{bmatrix} \in \mathbb{R}^{m\times n}$ with $\Sigma = \begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{bmatrix}$ (we know these n singlular values are strictly positive, since the rank of the matrix is n). We'd like to find a SVD for $A^T A \in \mathbb{R}^{n\times n}$.

By the transpose rules, $A^T = (U\Sigma V)^T = V^T \Sigma^T U^T$. Then we have:

$$\begin{aligned}
A^T A &= (V^T \Sigma^T U^T)(U\Sigma V) && \text{Above note} \\
&= V^{-1}\Sigma^T (U^{-1}U)\Sigma V && U \text{ and } V \text{ are orthogonal, plus associativity} \\
&= V^{-1} \begin{bmatrix}\Sigma^T & 0\end{bmatrix} \begin{bmatrix}\Sigma \\ 0\end{bmatrix} V && \text{The form } \Sigma \text{ takes} \\
&= V^{-1} Z^2 V && \text{(Block) matrix multiplication}
\end{aligned}$$

We claim $V^{-1}Z^2V$ is a valid SVD for $A^T A \in \mathbb{R}^{n\times n}$. To verify this, we check the three required properties. First, $V \in \mathbb{R}^{n\times n}$ is orthogonal from the decomposition of A. Second, $V^{-1} \in \mathbb{R}^{n\times n}$ is also orthogonal (we have $(V^{-1})^{-1} = (V^{-1})^T$ since $(V^{-1})^{-1} = V$ by the inverse rules), and since $(V^{-1})^T = (V^T)^T = V$ as well by the orthogonality of V. Finally, $Z^2 \in \mathbb{R}^{n\times n}$ is diagonal with non-negative entries from the decomposition of A. Note that this shows the singlular values of $A^T A$ are the square of the singlular values of A (see Problem 4.3).

Problem 4.2) Let $A \in \mathbb{R}^{m \times n}$ with $m > n$ and $\mathrm{Dim}(C(A)) = n$. Use a singlular value decomposition (SVD) of A to determine a SVD of AA^T.

Consider the SVD $A = U\Sigma V$ where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, and $\Sigma = \begin{bmatrix}\Sigma \\ 0\end{bmatrix} \in \mathbb{R}^{m\times n}$ with $\Sigma = \begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{bmatrix}$ (we know these n singlular values are strictly positive, since the rank of the matrix is n). We'd like to find a SVD for $AA^T \in \mathbb{R}^{m\times m}$.

By the transpose rules, $A^T = (U\Sigma V)^T = V^T\Sigma^T U^T$. Then we have:

$$\begin{aligned}
AA^T &= (U\Sigma V)(V^T\Sigma^T U^T) && \text{Above note} \\
&= U\Sigma(VV^{-1})\Sigma^T U^{-1} && U \text{ and } V \text{ are orthogonal, plus associativity} \\
&= U\Sigma\Sigma^T U^{-1} && \text{Inverse rules} \\
&= U \begin{bmatrix} Z \\ 0 \end{bmatrix} \; [Z \;\; 0] \; U^{-1} && \text{The form } \Sigma \text{ takes} \\
&= U \begin{bmatrix} Z^2 & 0 \\ 0 & 0 \end{bmatrix} U^{-1} && \text{(Block) matrix multiplication}
\end{aligned}$$

We claim $U(\Sigma\Sigma^T)U^{-1}$ is a valid SVD for $AA^T \in \mathbb{R}^{m\times m}$. To verify this, we check the three required properties. First, $U \in \mathbb{R}^{m\times m}$ is orthogonal from the decomposition of A. Second, $U^{-1} \in \mathbb{R}^{m\times m}$ is also orthogonal (we have $(U^{-1})^{-1} = (U^{-1})^T$ since $(U^{-1})^{-1} = U$ by the inverse rules, and since $(U^{-1})^T = (U^T)^T = U$ as well by the orthogonality of U and the transpose rules). Finally, $(\Sigma\Sigma^T) \in \mathbb{R}^{m\times m}$ is diagonal with non-negative entries from the decomposition of A. The block matrix multiplication above shows the first n singular values of AA^T are the square of the singular values of A, and the subsequent $m-n$ singlular values are zero (see Problem 4.3).

Problem 4.3) Let $A \in \mathbb{R}^{m \times n}$ with $m > n$ and $\mathrm{Dim}(C(A)) = n$. Express the individual singlular values of A^TA and AA^T in terms of those of A.

From Problem 4.1, we have the SVD $A^TA = V^{-1}Z^2V$ and so the singlular values of A^TA are the square of the singlular values of A.

From Problem 4.2, we have the SVD $AA^T = U(\Sigma\Sigma^T)U^{-1} = U \begin{bmatrix} Z^2 & 0 \\ 0 & 0 \end{bmatrix} U^{-1}$ and the first n singlular values of AA^T are the square of the singlular values of A, and the subsequent $m-n$ singlular values are zero.

Problem 4.4) Let $A \in \mathbb{R}^{m \times n}$ with $m > n$ and $\text{Dim}(C(A)) = n$. Use a SVD of A to determine a SVD of $(A^TA)^{-1}A^T$ and its individual singular values.

Consider the SVD $A = U\Sigma V^T$ where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, and $\Sigma = \begin{bmatrix} Z \\ 0 \end{bmatrix} \in \mathbb{R}^{m\times n}$ with $Z = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$ (we know these n singular values are strictly positive, since the rank of the matrix is n). We'd like to find a SVD for $(A^TA)^{-1}A^T \in \mathbb{R}^{n\times m}$.

By the transpose rules, $A^T = (U\Sigma V)^T = V^T\Sigma^T U^T$. From the derivation in Problem 4.1, $A^TA = V^{-1}Z^2V$. Then we have:

$$\begin{aligned}
(A^TA)^{-1}A^T &= (V^{-1}Z^2V)^{-1}(V^T\Sigma^TU^T) && \text{Above Notes}\\
&= (V^{-1}(Z^2)^{-1}V)(V^T\Sigma^TU^T) && \text{Inverse Rules}\\
&= (V^{-1}(Z^2)^{-1}V^{-1})(V^{-1}\Sigma^TU^{-1}) && \text{U and V are orthogonal}\\
&= V^{-1}(Z^2)^{-1}(VV^{-1})\Sigma^TU^{-1} && \text{Associativity}\\
&= V^{-1}((Z^2)^{-1}\Sigma^T)U^{-1}) && \text{Inverse rules}\\
&= V^{-1}\left(\begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{bmatrix}\begin{bmatrix} \sigma_1 & & & 0 & \cdots & 0 \\ & \ddots & & 0 & \cdots & 0 \\ & & \sigma_n & 0 & \cdots & 0 \end{bmatrix}\right)U^{-1} && Z\text{ Diagonal and form }\Sigma\text{ takes}\\
&= V^{-1}\begin{bmatrix} \frac{1}{\sigma_1} & & & 0 & \cdots & 0 \\ & \ddots & & 0 & \cdots & 0 \\ & & \frac{1}{\sigma_n} & 0 & \cdots & 0 \end{bmatrix}U^{-1} && \text{Matrix multiplication}
\end{aligned}$$

Where the second to last equality follows from the fact that the inverse of a diagonal matrix is the inverse of its diagonal elements, and the fact that Σ^T appends $m - n$ columns of zeros to Z.

We claim $V^{-1}([Z^{-1}0])U^{-1}$ is a SVD for $(A^TA)^{-1}A^T \in \mathbb{R}^{n\times m}$ (where $0 \in \mathbb{R}^{n\times(m-n)}$). To verify this, we check the three required properties. First, $V^{-1} \in \mathbb{R}^{n\times n}$ is orthogonal since $(V^{-1})^{-1} = V = (V^T)^T = (V^{-1})^T$ by the orthogonality of V. Second, $U^{-1} \in \mathbb{R}^{m\times m}$ is also orthogonal since $(U^{-1})^{-1} = U = (U^T)^T = (U^{-1})^T$ by the orthogonality of U. Finally, $[Z^{-1}0] \in \mathbb{R}^{n\times m}$ has a first block that is diagonal with non-negative entries from the decomposition of A (recall that all n singular values of A are strictly positive, and thus have an inverse, since A was full rank), and a final block of all zeros.

The form of $[Z^{-1}0]$ tells us that the singular values of $(A^TA)^{-1}A^T$ are the inverse of the singular values for A.

5 Projections And Least Squares

Problem 5.1) Let $\boldsymbol{A} = \mathbf{1_{17}}$ be the 17 $\times$ 1 vector of all ones, and $\boldsymbol{b}$ be the first column of the 17 $\times$ 17 identity matrix. Determine the solution $\widehat{\boldsymbol{x}}$ of the least squares problem $\min_{\boldsymbol{x}} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$.

The normal equation is $A^TAx = A^Tb$. Note that A^TA is a (non-zero) scalar and thus invertible, so there is a unique solution to the least-squares problem. It is:

$$\begin{aligned}
\widehat{x} &= (A^TA)^{-1}A^Tb && \text{Isolating } x \text{ from normal equation} \\
&= \frac{1}{17}A^Tb && A^TA = \sum_{i=1}^{17} A_i^2 = \sum_{i=1}^{17} 1^2 = 17 \\
&= \begin{bmatrix} \frac{1}{17} & \frac{1}{17} & \cdots & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} && \text{Form } A^T \text{ and } b \text{ take} \\
&= \frac{1}{17} && \text{Matrix multiplication}
\end{aligned}$$

Problem 5.2) Let $\boldsymbol{P} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$ be an orthogonal projector. Show that Range($\boldsymbol{P}$) = Null($\boldsymbol{I_n} - \boldsymbol{P}$).

We just use the idempotent property of orthogonal projectors, that $P^2 = P$. To show the desired result, we try for dual containment.

First, assume $a \in \text{Null}(I_n - P) = \{x \in \mathbb{R}^n : (I_n - P)x = 0\}$. We know such a vector exists since linear transformations always map zero to zero vectors. Then by distributivity, $(I_n - P)a = 0 \Rightarrow a - Pa = 0 \Rightarrow Pa = a$, and so a is in the range of P; we see $\text{Null}(I_n - P) \subset \text{Range}(P)$.

Next, assume $b \in \text{Range}(P) = \{y \in \mathbb{R}^n : \exists x \in \mathbb{R}^n \text{ s.t. } Px = y\}$. Again, we know such a vector exists because zero is an element of every vector space, and zero will always map to zero under linear transformations. Then by definition, $Px = b$ for some x. Since P is idempotent, we also have $P^2x = P(Px) = Pb$. Plugging the last equality into the second expression, we have $P(Px) = Pb = b$ and so $b - Pb = 0$; b is in the null space of $I_n - P$; $\text{Range}(P) \subset \text{Null}(I_n - P)$. This proves our result.

Problem 5.3) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ have orthonormal columns. Show that $\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{T}$.

Recall $A^{\dagger} = (A^TA)^{-1}A^T$. Since A has orthonormal columns, $(A^TA) = I_n$. It is then straightforward to see that $A^{\dagger} = (A^TA)^{-1}A^T = (I_n)^{-1}A^T = I_nA^T = A^T$.

Problem 5.4) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ have orthonormal columns, and let $\boldsymbol{b} = (\boldsymbol{I_m} - \boldsymbol{AA^T})\boldsymbol{e_n}$. Determine the solution $\widehat{\boldsymbol{x}}$ to the least squares problem $\min_{\boldsymbol{x}} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$.

Note that since A is orthonormal, $A^TA = I_n$, and is thus invertible; we will have a unique solution.

From the normal equation, we have:

$$
\begin{aligned}
\widehat{x} &= (A^TA)^{-1}A^Tb && \text{Definition}\\
&= I_n(A^Tb) && A \text{ has orthonormal columns}\\
&= A^T(I_m - AA^T)e_n && \text{How } b \text{ was defined}\\
&= (A^T - A^TAA^T)e_n && \text{Matrix multiplication is distributive}\\
&= (A^T - (A^TA)A^T)e_n && \text{Matrix multiplication is associative}\\
&= (A^T - I_nA^T)e_n && A \text{ has orthonormal columns}\\
&= 0e_n = 0 \in \mathbb{R}^n && \text{The } n\text{-dimensional vector with } n\text{-many zeros}
\end{aligned}
$$