Worksheet 3: Matrix-Vector Products (§1.3)

Problem 1: Vectors and Matrices.

Let us TEMPORARILY think of a "vector" as an element of $\mathbb{R}^n$, for some n. If $v \in \mathbb{R}^n$, we will say that v is an "n-vector." And let us continue, for now, to write vectors in columns. So, for instance, a 4-vector looks like this:

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \tag{$*$}$$

Eventually we will give a more general definition of "vector" and we will have examples of vectors that do not belong to any $\mathbb{R}^n$, but this is a good start.

(a) Explain how *addition* of vectors works, and explain any relevant restrictions (i.e., can *any* two vectors be added together?)

(b) Explain how *scalar multiplication* of vectors by scalars works, and explain any relevant restrictions (i.e., can *any* vector be scaled by *any* real number? By the way, what is a "scalar"?)

Solution:

(a) Two column vectors can be added together if and only if they have the same number of components, in which case addition is performed "component-wise," like this:

$$\vec{a} + \vec{b} \;=\; \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \;=\; \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix}.$$

(b) Any column vector can be multiplied by any scalar, and the scalar multiplication is performed "componenent-wise," like this:

$$c\vec{a} \;=\; c\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \;=\; \begin{bmatrix} ca_1 \\ \vdots \\ ca_n \end{bmatrix}.$$

Combining the operations of vector addition and scalar multiplication yields a very important construction called a "linear combination." If $\vec{v}_1, \ldots, \vec{v}_k$ is a list of vectors in $\mathbb{R}^n$ and if $c_1, \ldots, c_k$ is a list of scalars, then the sum

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$$

is called a *linear combination* of the vectors $\vec{v}_1, \ldots, \vec{v}_k$ with *weights*, or *coefficients* $c_1, \ldots, c_k$. Linear combinations will show up repeatedly throughout the course. A linear combination is *trivial* if every coefficient is zero, in which case the sum will be the zero vector.

A *matrix* is an array of numbers, like this:

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \\ 8 & 9 & 0 & 1 \end{bmatrix}$$

The *size* of a matrix is the number of rows and columns it has; this is expressed by saying that a matrix is "$m \times n$" (read: "m by n") if it has m rows and n columns. So the above matrix has size 3×4. The (i, j)-entry of a matrix A is the number in the ith row and jth column, and can be written $A(i, j)$. The matrix with (i, j)-entry a_{ij} is written $[a_{ij}]$.

(c) Explain how *addition* of matrices works, and explain any relevant restrictions (i.e., can *any* two matrices be added together?)

(d) Explain how *scalar multiplication* of matrices by scalars works, and explain any relevant restrictions (i.e., can *any* matrix be scaled by *any* real number?)

Solution:

(c) Two matrices can be added together if and only if they have the same size, in which case addition is performed "entry-wise," like this:

$$A + B = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{bmatrix}.$$

(d) Any matrx can be multiplied by any scalar, and the scalar multiplication is performed "entry-wise," like this:

$$cA = c\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & \cdots & ca_{1n} \\ \vdots & & \vdots \\ ca_{m1} & \cdots & ca_{mn} \end{bmatrix}.$$

Notice that if we write vectors as columns like in $(*)$ above, then vectors are special cases of matrices; that is, a vector $\vec{v} \in \mathbb{R}^n$ can be thought of as an $n \times 1$ matrix. This will occasionally be a useful point of view.

(e) What is a "zero matrix"? How many zero matrices are there? Note that if O is the $m \times n$ zero matrix and A is any $m \times n$ matrix, then $A + O = O + A = A$.

(f) If A is an $m \times n$ matrix, what is $-A$? What is $A + (-A)$? Does "subtraction" of matrices make sense, so that we can write $A - B$ instead of $A + (-B)$?

(g) Is matrix addition commutative? Associative? (If necessary, remind yourself what these terms mean!)

Solution:

(e) A *zero matrix* is just a matrix for which every entry is equal to zero. There are infinitely many zero matrices, exactly one for every possible size of matrix.

(f) If A is an $m \times n$ matrix, then $-A$ is just $(-1)A$, i.e., the matrix A scaled by -1. The matrix $-A$ is the "additive inverse" of A, since $A + (-A) = (-A) + A = O$ where O is the zero matrix of the same size as A. Subtraction of matrices makes good sense (as long as the two matrices have the same size!)

(g) Matrix addition is commutative, meaning $A + B = B + A$ whenever A and B have the same size, and matrix addition is also associative, meaning that $(A+B)+C = A+(B+C)$ whenever these sums are defined.

Problem 2: Linear Systems and Matrix-Vector Products.

In this problem we will start off easy, gradually complicate things, and then try to simplify them again to end up back where we started. A *linear equation in one variable* is an equation of the form

$$ax = b \tag{$\dagger$}$$

where a and b are real numbers and x is a variable. You have known how to solve such equations since grade school, but now we will increase the number of variables and the number of equations.

(a) Write the general form of a linear equation in *two* variables. $ax + by = c$

(b) Write the general form of a linear equation in n variables. $a_1x_1 + \cdots + a_nx_n = b$

(c) What type of object is a solution of a linear equation in n variables? An n-dimensional vector; i.e., an element of $\mathbb{R}^n$.

(d) How many solutions can a linear equation in n variables have? What is the geometric shape of the solution set of a linear equation in n variables? How do your answers depend on n? It has zero or infinitely many solutions, unless $n = 1$, in which case it has zero or one solution. The solution set, if nonempty, will be an $(n-1)$-dimensional hyperplane in $\mathbb{R}^n$.

(e) Write the general form of a *system of m linear equations in n variables.* Think carefully about how you want to name all the scalars and variables appearing in your equations.

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

(f) Now, rewrite your answer to part (e), which should have looked rather complicated, using matrices and vectors in such a way that it resembles the very simple equation ($\dagger$) above.

$$A\vec{x} = \vec{b}, \qquad \text{where} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

(g) If you did part (f) correctly, you should now be able to define *matrix-vector products*. That is, supposing $A = [a_{ij}]$ is a matrix and $\vec{x}$ is a (column) vector, give a definition of

$$A\vec{x},$$

making sure to indicate any restrictions on when $A\vec{x}$ is actually defined!
If A is $m \times n$ and $\vec{x} \in \mathbb{R}^n$, then $A\vec{x}$ is the vector in $\mathbb{R}^m$ whose ith component is $\sum_{j=1}^{n} a_{ij}x_j$, i.e., the dot product of the ith row of A with $\vec{x}$. If the number of components in $\vec{x}$ is different from the number of columns in A, then $A\vec{x}$ is not defined.

(h) Describe the matrix-vector product $A\vec{x}$ using an appropriate linear combination.
If the number of components in $\vec{x}$ is equal to the number of columns of A, then $A\vec{x}$ is the linear combination of the columns of A that results from using the components of $\vec{x}$ as weights.

(i) Express the following linear system as a matrix-vector product of the form $A\vec{x} = \vec{b}$.

$$\begin{array}{rcrcrcrcl} 0x_1 & + & 1x_2 & + & 2x_3 & + & 3x_4 & = & 1 \\ 4x_1 & + & 5x_2 & + & 6x_3 & + & 7x_4 & = & 2 \\ 8x_1 & + & 9x_2 & + & 0x_3 & + & 1x_4 & = & 3 \end{array}$$

Let $A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \\ 8 & 9 & 0 & 1 \end{bmatrix}$, let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$, and let $\vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

(j) Now that you have defined the matrix-vector product $A\vec{x}$, explain how you can understand the dot product of two vectors in $\mathbb{R}^n$ in terms of matrix-vector products. [Hint: this is our first example of an instance where it is useful to treat vectors sometimes as *rows* instead of always as columns!] If we write $\vec{x}$ as a row vector to the left of $\vec{y}$ written as a column vector and then multiply them according to the rules for matrix multiplication, the result will be a 1×1 matrix whose entry is $\vec{x} \cdot \vec{y}$.

(k) The *transpose* of a matrix A is written $A^\top$. Remind yourself what the transpose of a matrix is, and, using your answer to (j), explain why it is true that for all column vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$,

$$\vec{v}^\top \vec{w} \;=\; \vec{w}^\top \vec{v}.$$

By part (j), if $\vec{x}, \vec{y} \in \mathbb{R}^n$ then $\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y}$. Since the dot product is commutative, we therefore have

$$\vec{v}^T \vec{w} \;=\; \vec{v} \cdot \vec{w} \;=\; \vec{w} \cdot \vec{v} \;=\; \vec{w}^T \vec{v}.$$

Problem 3: Matrix-Vector Products and Functions.

Let X and Y be sets. A *function f from X to Y* is a rule that assigns to each element $x \in X$ a unique element $f(x)$ in Y, called the *value* of x under f. The set X is called the *domain*, or *source space* of f, and the set Y is called the *codomain*, or *target space* of f. The set

$$\mathrm{im}(f) \;=\; \{f(x) : x \in X\}$$

of all values of f is called the *image* of f. We write $f : X \to Y$ to indicate that f is a function from X to Y.

(a) Using matrix-vector products, explain how an $n \times m$ matrix can be viewed as a function from $\mathbb{R}^m$ to $\mathbb{R}^n$. In other words, given an $n \times m$ matrix A, define a function $f_A : \mathbb{R}^m \to \mathbb{R}^n$ in a natural way using matrix-vector products.

(b) A function f from $\mathbb{R}^m$ to $\mathbb{R}^n$ is *linear*[‡] if for all vectors $\vec{v}, \vec{w} \in \mathbb{R}^m$ and for all scalars $c \in \mathbb{R}$, the following hold:

$$\begin{aligned} f(\vec{v} + \vec{w}) &= f(\vec{v}) + f(\vec{w}); \\ f(c\vec{v}) &= cf(\vec{v}). \end{aligned}$$

For which matrices A is the function f_A that you defined in (a) above linear?

Solution:

(a) Given an $n \times m$ matrix A, define the corresponding function $f_A : \mathbb{R}^m \to \mathbb{R}^n$ by the rule $f_A(\vec{x}) = A\vec{x}$.

(b) Since matrix multiplication obeys the rules $A(B + C) = AB + AC$ and $A(cB) = cAB$ for all $c \in \mathbb{R}$ and matrices A, B, C for which the given products are defined, the function f_A will be linear for *every* matrix A.

[‡]This definition of *linear* generalizes the one from Worksheet 2, and is the definition we will use from now on throughout the course.