# Math 54 Course Summary

Matthew Signorotti, UC Berkeley

June–August 2018

# Contents


**1 Linear algebra** **2**
- 1.1 Systems of linear equations . . . 2
  - 1.1.1 Matrices . . . 3
  - 1.1.2 Solving systems of linear equations: Gaussian elimination . . . 4
  - 1.1.3 Two fundamental questions: existence and uniqueness of solutions . . . 7
- 1.2 Matrix algebra . . . 8
  - 1.2.1 The matrix of a linear transformation . . . 8
  - 1.2.2 Matrix multiplication . . . 9
- 1.3 Determinants (of square matrices) . . . 11
- 1.4 Orthogonality and least squares . . . 13
  - 1.4.1 . . . 13
  - 1.4.2 The least squares approximation . . . 13
- 1.5 Symmetric matrices . . . 14

**2 Differential equations** **15**
- 2.1 Preliminary definitions . . . 15
- 2.2 Solving second-order homogeneous differential equations . . . 15

**3 Glossary and Theorem Reference** **16**
- 3.1 Glossary . . . 16
- 3.2 Theorems . . . 17

# Chapter 1

# Linear algebra

## 1.1 Systems of linear equations

A **linear equation** is an equation of the form

$$c_1x_1 + c_2x_2 \cdots c_nx_n = m,$$

where $x_1, \ldots x_n$ are unknown values and $c_1, \ldots c_n$ and $m$ are known constants. The above expression left of the equals sign is called a **linear combination**: a sum of elements in a set, in this case the unknown values $x_1$ through $x_n$, each multiplied by some constant. Additionally, a **system of linear equations** is a group of linear equations whose solution(s) are all of the combinations of $x_1, \ldots x_n$ which satisfy every equation in the system.

Many problems in the real world can be analyzed in terms of systems of equations. Some problems involve fundamental questions of systems of equations including whether a solution exists, what such a solution might be, and if a solution is unique; all of these questions we will soon consider. The following are examples of problems involving systems of linear equations:

- **Circuit analysis**: For physical reasons, a circuit with $n$ elements including ideal wires, resistors, voltage sources, current sources, and open circuits is governed by linear equations with current and changes in voltage as the unknowns in a system of equations. Solving such a system reveals the equilibrium currents and voltage differences of the circuit, which are essential to how the circuit will interface with other electronics and whether it will safely work within the physical limitations of circuit elements.

- **Dieting**: Scientists in the 1908s constructed the ideal Cambridge diet for weight loss and healthy eating by using a system of equations. Since there is an ideal amount of each nutrient to be consumed daily, if we consider $n$ different foods a person might eat, then for each nutrient we have a linear equation

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n = m.$$

Here, $m$ is the ideal amount of the nutrient to be consumed daily, $c_i$ the amount of that nutrient obtained from the $i$th food, and $x_i$ the variable representing how much of the $i$th food to eat.

### 1.1.1 Matrices

Before we go any further into problems involving systems of linear equations, let's pose an equivalent but more convenient notation for such problems. Larger systems of equations can be a pain to write, so there are abbreviated are written as **matrix equations**. A **matrix** is an array of numbers which associates each number to a row and column. The following system of equations and two matrix equations are all equivalent. The third (and most abbreviated) style below is called an **augmented matrix** and is perhaps the most common, and is merely an abbreviation of the second style.

$$\begin{cases} x_1 + 2x_2 = 0 \\ -5x_1 - x_2 = 3 \end{cases} \qquad \begin{bmatrix} 1 & 2 \\ -5 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \qquad \left[\begin{array}{cc|c} 1 & 2 & 0 \\ -5 & -1 & 3 \end{array}\right]$$

A **vector** in $\mathbb{R}^n$ is an ordered list of $n$ values. In linear algebra, it is typically written as a matrix with one column; an example is the vector containing 0 and 3 in the second form above. Properties including vector-scalar multiplication, vector addition, and vector equality are defined just as you might expect. Before we go any further, let's formalize the concept of matrix multiplication with some definitions that justify why the above system of equation and matrix equations are equivalent.

**Definition 1** (matrix-vector multiplication)**.** *The matrix equation $A\mathbf{x}$ can be viewed as a linear combination of the columns of $A$ according to the weights given in $\mathbf{x}$. In other words, for $A = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \end{bmatrix}$ and $\mathbf{x}$ a vector in $\mathbb{R}^n$,*

$$A\mathbf{x} = \mathbf{a_1}x_1 + \mathbf{a_2}x_2 + \cdots + \mathbf{a_n}x_n.$$

This definition fits naturally with linear systems of equations, because we can represent any system of equations by populating the $i$th row of $A$ with the coefficients of the $i$th equation, treating $\mathbf{x}$ as the vector of unknowns, and setting $A\mathbf{x} = \mathbf{b}$, with $\mathbf{b}$ containing the constants each equation is equal to (see above for an example). The above form $A\mathbf{x}$ can be simplified to a vector containing the left side of the equals sign for all the equations in a system of equations we a representing as $A\mathbf{x} = \mathbf{b}$, and thus the matrix $A$ as described above has the same solution set as the system of equations with which we are concerned.

**Theorem 1.** *The matrix equation $A\mathbf{x} = \mathbf{b}$, defined as described above, has the same solution set as the system of equations it represents and the "vector equation" $\mathbf{a_1}x_1 + \mathbf{a_2}x_2 + \cdots + \mathbf{a_n}x_n = \mathbf{b}$.*

Ultimately, this theorem is justified by the rules we have built ourselves for matrix multiplication, particularly the above definition. By following logic relating to the conventions of vector addition, vector-scalar multiplication, and matrix-vector multiplication, one can go both ways between systems of equations and their equivalent matrix and vector equation representations. However, this theorem implies that we can phrase problems in many ways, using these three equation forms equivalently.

**Definition 2** (the row-vector rule)**.** *A notable implication of the above definition is that the ith row of the vector* $A\mathbf{x}$ *will be* $\sum_j a_{ij}x_j$*, where* $a_{ij}$ *is the entry in row i and column j, and j ranges between 1 and the number of entries in* $\mathbf{x}$*. In terms we will later come to understand, the ith row of* $A\mathbf{x}$ *is the dot product of the ith row vector of A with* $\mathbf{x}$*.*

By phrasing linear equations and linear transformations in terms of matrix equations and matrix-vector multiplication, we have unlocked new ways of describing and solving problems, as we will soon see. An example of an interesting property easily described by matrix multiplication is the **span** of a matrix's column vectors, which is all vectors reachable by $A\mathbf{x}$ for any $\mathbf{x}$.

Additionally, one might imagine modelling some sort of transformation or function as a matrix multiplication, since matrix multiplication takes in one vector and outputs another. This intuition turns out to be true, as we will see; any linear transformation can be modelled as multiplication by a matrix full of constant real numbers.

### 1.1.2 Solving systems of linear equations: Gaussian elimination

Perhaps the simplest problem in linear algebra is solving a system of equations for the exact solution(s) of that system. An algorithm known as **Gaussian elimination** or **row reduction** addresses this problem. It is based on three moves done to systems of equations:

1. **Interchanging** two equations, or switching the order in which they are presented
2. **Scaling** an equation by a nonzero real number
3. **Replacement** of one equation with the sum of that equation and a scalar multiple of another

"Solving" a system of equations also demands a method of expressing the solutions of a system when there are too many solutions to merely list out. It turns out that Gaussian elimination will produce a special form of matrix equations or systems of equations called **reduced row echelon form**, which intuitively defines some quantities in terms of only constants and other "free" variables which are allowed to vary.

#### The algorithm

Essentially, Gaussian elimination aims to keep eliminating variables from equations until the form of the solution is obvious — either no solution, one solution, or many solutions. The algorithm is as follows. Gaussian elimination is always performed on a matrix equation, but this is just a notation, since the underlying principles involve systems of equations.

1. Bring the matrix to **echelon form**, otherwise known as upper-triangular form. This form is defined by the following properties: (1) rows of all 0s are at the bottom, (2) the leading nonzero entry, or **pivot**, of each row is to the right of all pivots above it, and (3) all entries below a pivot are 0s.

- When a matrix is in this form, there will be a staircase of 0s going up and to the left, hence the name "upper triangular" form.
- This process can be described more algorithmically: Iterate left to right through the columns/unknown variables. Let $i$ be how many rows you have swapped to the top of the matrix. For iteration, swap with row $i + 1$ a row at or below row $i + 1$ which has a leftmost pivot. Then, use row addition (and potentially scaling for ease of row addition) to eliminate the unknown variable corresponding to the pivot of the row you just swapped from all equations below row $i + 1$. Increment $i$ and repeat until there are no more rows to swap.

2. From right to left, use row replacement to create 0s above every pivot. At any point in this process, scale each row to have a pivot of 1. A matrix in echelon form also with 0s above each pivot and each pivot being 1 is reduced echelon form, as mentioned above.

3. Finally, we will have reached the ideal state to read the solution set from, reduced echelon form. The solution set of these equations, for reasons presented in the next section, is equivalent to the solution set of the original equations.

   - We may have an equation such as $0 \neq 1$ equating 0 to a nonzero constant. No matter what values we give the unknowns, this will never be true, so the solution set is empty and the system of equations is called **inconsistent**, or having no solution.
   - If we don't have such an equation, then every equation equates a linear combination of the unknown variables to a constant. By subtracting out all variables except the first (pivot) variable of an equation, we can define each pivot variable in terms of a constant and a linear combination of non-pivot variables, which are called **free variables**.
     - Note that because the matrix is in reduced echelon form, each equation's pivot variable is *not* at all a term in any of the other equations. Therefore, we have defined each of the pivot variables in terms of constants and the free variables.

Now, we have an equation for every pivot variable. Perhaps there are no free variables and all of the pivots have only one value; in this case, there is only the one solution. If not, we can clearly freely choose values for the free variables, in terms of which the pivot variables are defined, and these values will determine the values of the pivot variables for a solution.

**Parametric vector form** is a convenient way to express the solution set when there is one or more free variables. Parametric vector form simply expresses the solution as a constant vector plus a vector times each free variable. For instance, the following solutions are equivalent, with the right solution in parametric form:

$$\mathbf{x} = \begin{bmatrix} -6 - x_2 + x_3 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} -6 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

**Proof of the Gaussian elimination method**

**Theorem 2.** *Any system of equations reached through the three basic row operations of Gaussian elimination will have the same solution set as the original system of equations. In other words, all the matrices seen during row reduction are* ***row equivalent****, or have the same solution set.*

*Proof.* In Gaussian elimination, a matrix's corresponding equations are manipulated through only the three basic row operations. I propose that we examine why, if the equations of concern are all fulfilled by some solution before each of the operations, the equations will be true after. Answering this question boils down to self-evident mathematical axioms such as those presented in Euclid's "Common Notions." In the following list of the three operations, I will consider why a solution to the system of equations before each operation, or move, will satisfy all the equations after the move, citing axioms in the process. Additionally, I will show that each move is reversible, in that you can return to the original set of equations even after performing one of the row operations. Later on, I will use the reversible property of these operations to show that Gaussian elimination to argue that the solution set is unchanged.

1. Interchange: This row operation does not change the solution set to the equations because the equations themselves are unchanged, and a potential solution will still need to satisfy all of the equations. This move is easily reversible; simply switch the rows back. Interchange is primarily just to track which rows we have swapped to the top while approaching upper triangular form.

2. Scaling: It is widely accepted that two equals scaled by the same constant will be equal. One could argue this is because scalar multiplication is a function, and a basic axiom of equality says that equal quantities used as equal inputs for a function result in equal outputs. Scaling is reversible because, after multiplying an equation by $c \neq 0$, one can multiply it by $1/c$ to obtain the original equation.

3. Replacement: In an addition or replacement, a multiple of one row is added to another row. One of Euclid's axioms observed that "If equals are added to equals, then the wholes are equal." If both equations were satisfied before, then the new equation will be equal and satisfied as well. Additionally, if in the original transformation $c$ times one equation was added to the other, the row operation can be reversed by adding $-c$ times that equation.

Now, for notation's sake, call the set of equations before a row operation $A$, and the set of equations after an operation $B$. The set of solutions to the equations of $A$ will be referred to as $A$'s "solution set." Due to the axioms and reasoning I provided above, the set of equations $B$ should still hold true for anything in $A$'s solution set. In other words, $B$'s solution set contains the entirety of $A$'s solution set.

Although this logic indicates that everything in $A$'s solution set is in $B$'s, how do we know the converse is true: that nothing is in $B$'s solution set but not $A$'s? Can $A$'s solution set grow when we advance from $A$ to $B$? To answer this question, recall that all of the row operations above are reversible. Consider using a row operation to go from $A$ to $B$ and then back to $A$. Following the same logic, we conclude that not only

will $B$ still hold true for $A$'s solution set, but $A$ will also hold true for $B$'s solution set. In other words,

(1) everything in $A$'s solution set is in $B$'s solution set, and

(2) everything in $B$'s solution set is in $A$'s.

At last, we are prepared to answer our motivating question: Is it possible that a solution will be unique only to $A$ or $B$'s solution set? A solution cannot just be in $A$ due to (1), and a solution cannot just be in $B$ because of (2). A solution being in only either solution set is impossible, so any potential solution will be in either both sets or neither one. Because they have the same contents, the solution sets of $A$ and $B$ are identical. Now, we have proven that the solution set of a system of equations does not change during any of the three above row operations. Extending this logic, we can conclude that performing any number of row operations will not alter the solution set, so Gaussian elimination is justified. □

### 1.1.3 Two fundamental questions: existence and uniqueness of solutions

**Question 1: the existence of a solution**

After performing Gaussian elimination, if there is a row such as "$0 = 1$" indicating an inconsistent system, then no solution exists. Otherwise, however, we will have definitions of the pivot variables in terms of constants and free variables which we can freely vary, so a solution will exist in this case.

More generally, does an $m \times n$ matrix, or set of vectors, **span** $\mathbb{R}^m$? (A matrix's columns being "spanning" means that $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$.) The following statements are equivalent:

- For each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution (stating the problem as a matrix equation).
- Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of $A$ (stating the problem as a vector equation).
- The columns of $A$ span $\mathbb{R}^m$.
- $A$ has a pivot position in every row.
  - Why must there be a pivot in every row? If there weren't, there would be a row of all 0s, and it would be possible to choose a $\mathbf{b}$ that leads to an inconsistent equation like $0 = -4$, which shows the system is inconsistent.

The above theorem is the foundation for the **invertible matrix theorem**; as we progress through linear algebra, many more equivalent conditions will be added to this list.

**Question 2: the uniqueness of a solution**

When is there just one unique solution? Recall that a free variable indicates an infinite number of solutions since we can freely choose any real number as its value. Therefore, a solution is unique if there is no free variable, of course provided a solution does exist. For there to be no free variable, every column must have a pivot. And as we established above, for a solution to exist, every row has a pivot. Therefore, there must be a pivot in every row and column. This condition occurs exactly when the matrix $A$ in question is square, or $n \times n$, and row reduces to the **identity matrix** $I$, which has 1s on the diagonal and 0s everywhere else.

A related question is when is there no solution *or* just one solution to a system. This time, we just need to make sure there are no free variables, so again we impose that a pivot is in every column. However, there need not be a pivot in every row; the system can be inconsistent and have no solution.

As a final point, a linear system of equations either has no solution, one solution, or infinitely many solutions. The following is why.

**Theorem 3.** *The existence of two solutions to a matrix equation implies the existence of an infinite number of solutions.*

*Proof.* Call the two solutions to $A\mathbf{x} = \mathbf{b}$ $\mathbf{u}$ and $\mathbf{v}$. We can choose an infinite number of two real numbers $a$ and $b$ such that

$$m + n = 1$$

Then, by the distributive and scalar-multiplication properties of matrix multiplication,

$$A(m\mathbf{u} + n\mathbf{v}) = mA\mathbf{u} + nA\mathbf{v} = (m + n)\mathbf{b} = \mathbf{b}$$

□

# 1.2 Matrix algebra

## 1.2.1 The matrix of a linear transformation

**Definition 3.** *A function $T$ is linear if and only if the following properties hold:*

$$T(x + y) = T(x) + T(y)$$
$$T(cx) = cT(x)$$

**Theorem 4.** *Any linear transformation $T(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ can be implemented with an $m \times n$ matrix, so that $T(\mathbf{x}) = A\mathbf{x}$. $A = \begin{bmatrix} T(\mathbf{e_1}) & T(\mathbf{e_2}) & \cdots & T(\mathbf{e_n}) \end{bmatrix}$, where $\mathbf{e_i} \in \mathbb{R}^n$ consists of all zeroes except 1 in the ith entry.*

This theorem allows for easy computation of linear transformations by our rules of matrix multiplication. For instance, consider rotation by $\theta$ in two dimensions. Geometrically, this operation is linear for vectors in the Cartesian plane. We just need to know the rotated versions of $(1, 0)$ and $(0, 1)$ to implement a rotation matrix.

*Proof.* First, observe that $\mathbf{x} = x_1\mathbf{e_1} + x_2\mathbf{e_2} + \cdots + x_n\mathbf{e_n}$. Therefore, $T(\mathbf{x}) = T(x_1\mathbf{e_1} + x_2\mathbf{e_2} + \cdots + x_n\mathbf{e_n})$, and we can exploit the linear properties of $T$ to further show that for any input $\mathbf{x}$, we can write

$$T(\mathbf{x}) = x_1T(\mathbf{e_1}) + x_2T(\mathbf{e_2}) + \cdots + x_nT(\mathbf{e_n}).$$

Now, we want a matrix such that $T(\mathbf{x}) = A\mathbf{x}$. In other words,

$$T(\mathbf{x}) = x_1\mathbf{a_1} + x_2\mathbf{a_2} + \cdots + x_n\mathbf{a_n} = x_1T(\mathbf{e_1}) + x_2T(\mathbf{e_2}) + \cdots + x_nT(\mathbf{e_n}).$$

As we can see, following our definition of matrix multiplication, and because the transformation can be interpreted linearly, $T(\mathbf{x}) = A\mathbf{x}$ always, provided that $\mathbf{a_i} = T(\mathbf{e_i})$. □

### 1.2.2 Matrix multiplication

Matrices can also describe any composition of linear functions $T_A \circ T_B(\mathbf{x}) = T_A(T_B(\mathbf{x}))$, because the composition of linear functions can be shown to be also linear. Matrix multiplication allows us to do just designed to accommodate composition of linear functions. Due to the properties of linearity, composing two linear functions results in a new linear function, so the matrix representation of $T_A(T_B(\mathbf{x}))$ is

$$\begin{bmatrix} T_A(T_B(\mathbf{e_1})) & T_A(T_B(\mathbf{e_2})) & \cdots & T_A(T_B(\mathbf{e_n})) \end{bmatrix}.$$

If we already have matrices $A$ and $B$ representing $T_A$ and $T_B$, then the above matrix becomes

$$\begin{bmatrix} T_A(B\mathbf{e_1}) & \cdots & T_A(B\mathbf{e_n}) \end{bmatrix} = \begin{bmatrix} T_A(\mathbf{b_1}) & \cdots & T_A(\mathbf{b_n}) \end{bmatrix} = \begin{bmatrix} A\mathbf{b_1} & \cdots & A\mathbf{b_n} \end{bmatrix}.$$

For convenience, we generalize this last matrix to be the definition of matrix multiplication.

**Definition 4** (matrix multiplication)**.** *The expression $AB$ describes the function composition $T_A \circ T_B$, where $T_A$ and $T_B$ are the transformations represented by A and B. By the above logic, we define*

$$AB = \begin{bmatrix} A\mathbf{b_1} & A\mathbf{b_2} & \cdots & A\mathbf{b_n} \end{bmatrix}.$$

From this definition, one can use the row-vector definition of matrix multiplication to extend the row-vector rule for matrix multiplication: The entry in row $i$ and column $j$ of $AB$ will be $\sum_j a_{ij}b_{i,j}$, where $a_{ij}$ is $A$'s entry in row $i$ and column $j$, and $b_{i,j}$ is the $j$th entry of $\mathbf{b_i}$. We will eventually learn that this is the same as the dot product of the $i$th row vector in $A$ with $\mathbf{b_j}$. Additionally, note that for such a function composition to carry meaning, the input dimension of $A$ must match the output dimension of $B$; this condition happens to be true when $A\mathbf{b_i}$ is defined for all $i \in [1, n]$.

#### Function inverses

*The associative property*: Our definition implies the associative property of matrix multiplication. In this case, the original operation resembled $A(B\mathbf{x})$, or more simply $AB\mathbf{x}$,

with $\mathbf{x}$ the input and $A$ and $B$ the composed linear functions. However, since we can prove that the resulting function composition is linear, this is equivalent to $(AB)\mathbf{x}$. Normally, in a long chain of matrix multiplications, we do each multiplication from right to left, starting with the rightmost matrix and the input vector — something of the form $A(B(C(D(E \cdots \mathbf{x}))))$. For any two consecutive matrices, we can model everything right of them as an input vector and thus do a function composition, grouping them together.

Note: Unlike multiplication in $R$, there is no commutative property for matrix multiplication. $AB \neq BA$ (with some exceptions, such as $A = B = I$).

### Scalar multiplication

Just how it sounds: multiply every entry of the matrix by a scalar.

### Matrix addition

Just how it sounds. $(A + B)\mathbf{x}$ models the addition of two linear transformations, given by $A$ and $B$.

### Matrix inverses

Function composition can be extended to derive an "inverse" function matrix, one which, when composed with the normal function, creates the identity function: $ToT^{-1}(\mathbf{e}_i) = \mathbf{e}_i$. We can imagine that if we were to to model $ToT^{-1}$ as a matrix multiplication, then $AA^{-1} = \begin{bmatrix} A\mathbf{a}_1^{-1} & A\mathbf{a}_2^{-1} & \cdots & A\mathbf{a}_n^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}$. The strategy to find the matrix inverse thus reduces to $n$ Gaussian elimination problems, with $A\mathbf{a}_i^{-1} = \mathbf{e}_i$. Rather than row reducing the same matrix $n$ times, we do all the problems simultaneously by putting all the $\mathbf{a}_i^{-1}$ together to the right, in a process called Gauss-Jordan elimination. More information on the algorithm is available with a quick Google search.

Keep in mind that the following statements regarding a square matrix are either all true or all false:

- A linear transformation $T$ has an inverse.
- The corresponding matrix $A_T$ has an inverse.
- $T$ is onto.
- $T$ is one-to-one.
- $A_T$ has an inverse.
- $T$ has an inverse.
- The columns of $A_T$ are linearly independent and span $\mathbb{R}^n$.
- $A_T$ has a pivot in each row and column.
- $A_T$ row reduces, or is row equivalent to, the identity matrix.

**Other properties**

1. $ABC = A(BC) = (AB)C$
2. $A(B + C) = AB + AC$
3. $(A + B)C = AC + BC$
4. $\lambda(AB) = (\lambda A)B = A(\lambda B)$
5. $I_m A = A = AI_n$ (where $A$ is $m \times n$ and $I_m$ is the square identity matrix, with ones down the diagonal and zeroes everywhere else)
6. $(AB)^{-1} = B^{-1}A^{-1}$

## 1.3 Determinants (of square matrices)

**Definition 5** (the determinant)**.** *For a square matrix $A = \left[a_{ij}\right]$ with at least 2 rows, the determinant of $A$ is defined as*

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j},$$

*where $A_{ij}$ is $A$ without the $i$th row and $j$th column, and $a_{ij}$ is the entry of $A$ in the $i$th row and $j$th column.*

**Theorem 5.** *The "**cofactor expansion**" of $A$ along any of $A$'s rows and columns is equal to the determinant of $A$. In particular, the cofactor expansion along the $i$th row is*

$$\sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} = \det A,$$

*and the cofactor expansion along the $j$th column is*

$$\sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} = \det A.$$

The following few known facts on determinants allow for the easy computation of determinants of huge square matrices. This method is much more efficient than the traditional alternative, which has very inefficient computational complexity. However, the traditional method does tell us the useful bit of information that the determinant of a square matrix in echelon form will be the product of all its diagonal entries, which will either all be pivots or contain 0.

1. Interchanging two rows negates the determinant.
2. Scaling a row by a constant multiplies the determinant by that constant.
3. Adding a scalar multiple of one row to another does not change the determinant.

These three moves are the three moves used in Gaussian elimination. Therefore, one can go about Gaussian elimination as normal until reduced row echelon form is reached. The determinant of the reduced row echelon form must either be one or zero because in this form, either the matrix is $I$, or it has a row of zeros. A more rigorous explanation of this process, along with useful formulas for calculating the determinant, can be found at Project Rhea.

#### Properties (left unproven)

- $\det AB = \det A \det B$
- $\det A^T = \det A$
- The determinant of an upper- or lower-triangular matrix is the product of the diagonal entries.

## Vector spaces

A vector space is a set in which addition and scalar multiplication are defined. Examples include not just $\mathbb{R}$ and $\mathbb{R}^m$, but

- $\{f : \mathbb{R} \to \mathbb{R}\}$
- Continuous functions $\{f : \mathbb{R} \to \mathbb{R}\}$
- Integrable functions $\{f : [a, b] \to \mathbb{R}\}$
- Polynomials with real coefficients
- Solutions to differential equations

Vector spaces have a number of additional, more specific properties, including associativity and commutativity of addition, the distributive property for scalar multiplication, and additive and multiplicative identity. "Scaling" and "addition" can be defined in any way, as long as the properties hold. Linear algebra concepts such as span and linear independence transfer over nicely as vector spaces, but row reduction makes no sense in terms of vector spaces.

#### Subspaces

A subspace is a subset of a vector space $V$ with the same properties of addition and multiplication, but in which (1) $\mathbf{0} \in U$, (2) $\mathbf{x}, \mathbf{y} \in U \longrightarrow \mathbf{x}+\mathbf{y} \in U$, and (3) $\mathbf{x} \in U \longrightarrow \lambda\mathbf{x} \in U$.

### Bases

A basis provides a coordinate system for a vector space. A basis maps from $R^n$ to $V$, even if $V$ cannot be expressed as $R^m$ for some $m$. A linear combination of basis vectors can form any vector in a vector space.

We can convert between bases for a vector space by manipulating the linearity of such a transformation. The $i$th column will be the transformation from $\mathbf{e}_i$ in b-coordinates to c-coordinates. $\mathbf{e}_i$ in b-coordinates represents the $i$th basis vector $\mathbf{b}_i$ in the vector space, by definition of a basis as a coordinate system. Now, we just need to convert from the vector space to c-coordinates. If the vector space is some form of $R^n$, this is a matter of Gaussian elimination, solving for $C[\mathbf{b}_i]_c = \mathbf{b}_i$. Therefore, for a vector space that is $R^n$, set up a Gauss-Jordan elimination problem with $C$ on the left half of the matrix and $B$ on the right half. However, for more complicated vector spaces such as parabolas, we will have to write each $i$th basis vector as a linear transformation of the c-coordinate basis vectors.

### Eigenvectors, values, and spaces

An eigenvector $\mathbf{x} \in R^n$ of an $n \times n$ matrix A satisfies (1)$\mathbf{x} \neq \mathbf{0}$ and (2) $A\mathbf{x} = \lambda\mathbf{x}$ for some $\lambda \in R$. The $\lambda$ is referred to as the "eigenvalue" corresponding to the eigenvector $\mathbf{x}$. Note that the span of eigenvectors corresponding to an eigenvalue form a vector space known as an eigenspace.

To find eigenvalues and from those their corresponding eigenvectors, we solve the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ (for a nontrivial solution). By the Invertible Matrix Theorem, this is equivalent to solving for the determinant of $A - \lambda I$ and finding which $\lambda \in R$ satisfy this determinant. The resulting polynomial involving $\lambda$ is called a "characteristic polynomial." The algebraic multiplicity of $a$ is the degree on $(\lambda - a)$ in the characteristic polynomial. The geometric multiplicity is the dimension of the eigenspace, and must be $\geq 1$ and $\leq$ the algebraic multiplicity.

## 1.4 Orthogonality and least squares

### 1.4.1

### 1.4.2 The least squares approximation

**Least-squares problems** concern finding a vector $\mathbf{x}$ such that $\|\mathbf{b} - A\mathbf{x}\|$ is minimized. (The name "least-squares" arises from the fact that $\|\mathbf{b} - A\mathbf{x}\|$ is the square root of a sum of squares.) Note that $A\mathbf{x}$ is always in Col$A$, so let's reduce the problem to finding the vector $\mathbf{v} \in \text{Col}A$ which minimizes $\|\mathbf{b} - \mathbf{v}\|$. The best approximation theorem tells us that for $A\hat{\mathbf{x}} = \hat{\mathbf{b}} = \text{proj}_{\text{Col}A}\mathbf{b}$, $\|\mathbf{b} - \hat{\mathbf{b}}\| \leq \|\mathbf{b} - \mathbf{v}\|$ for all $\mathbf{v} \in \text{Col}A$. And since $A\mathbf{x} \in \text{Col}A$, $\|\mathbf{b} - \hat{\mathbf{b}}\| = \|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$ for all $A\mathbf{x}$.

Through careful application of the best approximation theorem, and the fact that a projection onto the column space of $A$ is in Col$A$ and thus equals $A\hat{\mathbf{x}}$ for some $\hat{\mathbf{x}}$, we have shown that $\mathbf{x} = \hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ solves the original problem of minimizing $\|\mathbf{b} - A\mathbf{x}\|$. We also know that there is at least one valid $\hat{\mathbf{x}}$ solving $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, since $\hat{\mathbf{b}} \in \text{Col}A$.

One issue remains: actually finding a $\hat{\mathbf{x}}$. We can start off by applying the orthogonal decomposition theorem, which reveals that $\mathbf{b}-\hat{\mathbf{b}}$ is orthogonal to Col$A$ and thus $A^T(\mathbf{b}-\hat{\mathbf{b}}) = A^T(\mathbf{b}-A\hat{\mathbf{x}}) = \mathbf{0}$. Using the additive linearity of matrix transformations, we can manipulate this equation to get

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

Here we see what is essentially a classic matrix equation, with at least one solution. If we can show that $A^T A$ is invertible, then we can left-multiply by this inverse and get the one solution $\hat{\mathbf{x}}$. But if $A^T A$ is not invertible, its columns are linearly dependent by the invertible matrix theorem and thus there are infinite solutions.

**Theorem 6.** *Test*

## 1.5 Symmetric matrices

**Theorem 7.** *$A$ is symmetric $\implies$ any two eigenvalues from different eigenspaces of $A$ are orthogonal.*

# Chapter 2

# Differential equations

## 2.1 Preliminary definitions

A **differential equation** is an equation relating various derivatives of a function. The most common example is Newton's second law:

$$\mathbf{F} = m\frac{d^2\mathbf{x}}{d\mathrm{t}^2} \text{ (as a vector equation), or } F = m\frac{d^2\mathrm{x}}{d\mathrm{t}^2}.$$

A **homogeneous differential equation** equates an expression involving derivatives of a function to 0. For example, say we are examining an ideal oscillating spring in one dimension. A physical law states that the force of oscillation is negatively proportional to the displacement $x$ from the equilibrium position. Thus, Newton's second law from above becomes the following homogeneous differential equation:

$$F = -kx = m\frac{d^2\mathrm{x}}{d\mathrm{t}^2} \implies kx + m\frac{d^2\mathrm{x}}{d\mathrm{t}^2} = 0.$$

Lastly, an **nth-order differential equation** contains the term $y^{(n)}$ as the highest-order derivative in the expression. The above oscillating spring example involves a second-order differential equation because the expression's highest order of derivation is 2.

## 2.2 Solving second-order homogeneous differential equations

Any differential equation $ay'' + by' + cy = 0$ has a two-dimensional solution space. For any $t_0$, knowing $y(t_0)$ through $y_(n-1)(t_0)$ gives us just one possible function for $y$. In other words, the transformation from functions $y$ to $y(t_0)$ and $y'(t_0)$.

# Chapter 3

# Glossary and Theorem Reference

## 3.1 Glossary

- **Homogeneous** system of equations: a system of equations of the form $A\mathbf{x} = \mathbf{0}$
- **Nonhomogeneous** system of equations: a system of equations of the form $A\mathbf{x} = \mathbf{b} \neq \mathbf{0}$
- **Parametric vector form**: Writing the solution set as follows,

$$\mathbf{x} = \begin{bmatrix} 4 \\ -7 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

  with $x_3$ being a free variable. This contrasts with the traditional way of writing solutions:

$$x_1 = 4 + x_3$$

$$x_2 = -7 - x_3$$

$$x_3 \text{ is free}$$

- **Image**: a specific output of a function
- **Range**: the set of all images reachable by some function $T$
- **Domain**: the set of valid inputs
- **Codomain**: the dimension or type of output generated by a function $T$ — for instance, $\mathbb{R}^n$

- **Onto/surjective**: The range is the same as the codomain. A linear function $T$ is onto iff $T(\mathbf{x}) = A_T\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$, iff the columns of $A_T$ span $R_n$, or iff there is a pivot in every row. This property can be proved through row reduction.

- **Injective** (a.k.a. "one-to-one"): All inputs map to a unique output. For all images $y$, here are no distinct inputs $x_1$ and $x_2$ for which $f(x_1) = f(x_2)$. For a function in two dimensions to be injective, the horizontal line test must succeed (no horizontal lines intersecting two points). A linear function $T$ is injective iff $T(\mathbf{x}) = A_T\mathbf{x} = \mathbf{b}$ always has at most one solution, or iff the columns of $A_T$ are linearly independent i.e. a pivot is in every column. This property can be proved through row reduction.

## 3.2 Theorems

- Invertible matrix theorem:

- Orthogonal decomposition theorem: