Review: Basics of Linear Algebra

An Algorithmist's Toolkit (CSCI 2952T)

D Ellis Herskowitz
Brown University

The goal of this is to quickly remind you of the basics of linear algebra, some of which we will use in class. I assume that you have already taken at least one linear algebra class and are familiar with standard notation from linear algebra. I won't provide any proofs but if you're looking for more resources and/or proofs I strongly recommend *Linear Algebra Done Right* by Axler.

For each of the problems you are asked to solve, you are free to assume any of the preceding facts in the document (other than the fact itself).

1 Subspaces, Span, Independence, Bases and Dimension

We begin with some basic notions related to sets of vectors.

Definition 1.1 (Subspaces) *A set of vectors $V \subseteq \mathbb{R}^n$ is called a subspace if for any $u, v \in V$ and $a, b \in \mathbb{R}$ we have*

$$au + bv \in V.$$

Problem 1

If $V \subseteq \mathbb{R}^n$ is a subspace then 0 (the all zeros vector) is in V.

The span of a set of vectors is the result of all possible linear combinations of those vectors; namely, all possible ways of scaling and then adding together vectors in the set.

Definition 1.2 (Span) *Given a set of vectors $V = \{v_1, v_2, \ldots\} \subseteq \mathbb{R}^n$, the span of V is defined as*

$$\text{SPAN}(V) := \left\{ \sum_i a_i v_i : a_1, a_2, \ldots \in \mathbb{R} \right\}$$

Problem 2

The span of any set of vectors is a subspace.

Linear independence tells us whether or not one can reconstruct a vector in a set by linearly combining other vectors in that set.

Definition 1.3 (Linear Independence) *A set of vectors $V = \{v_1, v_2, \ldots\} \subseteq \mathbb{R}^n$ is said to be linearly dependent if there exist not all zero $a_1, a_2, \ldots \in \mathbb{R}$ such that*

$$\sum_i a_i v_i = 0.$$

Equivalently, V is dependent if there exist $a_1, a_2, \ldots \in \mathbb{R}$ and some $v_i \in V$ such that

$$v_i = \sum_{j \neq i} a_j v_j.$$

If V is not dependent then we say that it is linearly independent (independent for short).

An important class of independent sets of vectors are bases.

Definition 1.4 (Bases) *Given a set of vectors $V \subseteq \mathbb{R}^n$ and $B \subseteq V$, we say that B is a basis of V if*

$$\textsc{span}(B) = V \qquad \text{and} \qquad B \text{ is independent.}$$

An important fact is that every basis has the same size (i.e. number of vectors in it).

Fact 1.5 *Given a set of vectors $V \subseteq \mathbb{R}^n$ and bases B_1 and B_2 of V we have*

$$|B_1| = |B_2|.$$

A notion closely related to bases is that of dimension.

Definition 1.6 (Dimension) *The dimension of a subspace $V \subseteq \mathbb{R}^n$ is the size of the largest set of linearly independent vectors that it contains, namely*

$$\textsc{dim}(V) := \max_{U \subseteq V : U \text{ independent}} |U|.$$

An important fact is that the dimension of a set is always the size of any one of its bases. Using this fact, try and show the following.

Problem 3

$\textsc{dim}(\mathbb{R}^n) = n$.

Observe that the above means that this sense of dimension captures the usual sense for $\mathbb{R}^n$.

2 Linear Functions

The most important class of functions studied in linear algebra are so-called linear functions, defined as follows.

Definition 2.1 (Linear Functions) *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear if*

1. *$f(x + y) = f(x) + f(y)$ for any vectors $x, y \in \mathbb{R}^n$*
2. *$f(c \cdot x) = c \cdot f(x)$ for any vector $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$.*

Definition 2.2 (Range) *Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the range of f is*

$$\textsc{range}(f) := \{f(u) : u \in \mathbb{R}^n\}.$$

Notice that the range is a subset of $\mathbb{R}^m$.

Definition 2.3 (Kernel) *Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the kernel of f is*

$$\text{KER}(f) := \{u \in \mathbb{R}^n : f(u) = 0\}.$$

Notice that the kernel is a subset of $\mathbb{R}^n$

Problem 4

The range and kernel are both subspaces (and therefore both contain 0).

The following summarizes the intuitive idea that the parts of the input space ($\mathbb{R}^n$) that do not get mapped to interesting parts of the output space ($\mathbb{R}^m$) must be mapped to 0 in the output space.

Fact 2.4 (Fundamental Theorem of Linear Maps, $\mathbb{R}^n$) *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function then*

$$n = \text{DIM}(\text{KER}(f)) + \text{DIM}(\text{RANGE}(f)).$$

2.1 Matrices

Matrices occur in linear algebra because they give a natural and unique way of encoding linear functions. Specifically, every linear function corresponds to a (unique matrix) and vice versa.

Fact 2.5 *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function, then there exists a (unique) $m \times n$ matrix A s.t.*

$$f(x) = Ax$$

for every $x \in \mathbb{R}^n$.

Fact 2.6 *If A is an $m \times n$ matrix then there exists a (unique) linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ s.t.*

$$Ax = f(x)$$

for every $x \in \mathbb{R}^n$.

Given the correspondence between matrices and linear maps, we will often abuse notation and use a linear function and its matrix interchangeably. For example, we will talk about $\text{KER}(A)$ by which we really mean $\text{KER}(f)$ where f is the linear map corresponding to A.

For a given $m \times n$ matrix A, we let $\text{ROWS}(A) \subseteq \mathbb{R}^n$ be all rows of A treated as vectors in $\mathbb{R}^n$ and we let $\text{COLS}(A) \subseteq \mathbb{R}^m$ be all columns of A, treated as vectors in $\mathbb{R}^m$.

One nice way of thinking about what a matrix A with columns $c_1, c_2, \ldots c_n \in \mathbb{R}^m$ is doing as a function is it takes in a vector $x \in \mathbb{R}^n$ and then linearly combines its columns using x as the coefficients of this combination. That is,

$$Ax = \sum_{c_j \in \text{COLS}(A)} x_j c_j$$

where in this sum $x_j \in \mathbb{R}$ and $c_j \in \mathbb{R}^m$. Along these lines, check the following for yourself.

Problem 5

$\text{SPAN}(\text{COLS}(\text{A})) = \text{RANGE}(\text{A})$.

The following shows that the row and column spans of a matrix have equal dimension.

Fact 2.7 (Row Rank = Col Rank) *Given a matrix A, we have*

$$\text{DIM}(\text{SPAN}(\text{ROWS}(A))) = \text{DIM}(\text{SPAN}(\text{COLS}(A))).$$

Given the above equality, we can define the rank of a matrix as follows.

Definition 2.8 (Rank of a Matrix) *We let* $\text{RANK}(A) := \text{DIM}(\text{SPAN}(\text{ROWS}(A))) = \text{DIM}(\text{SPAN}(\text{COLS}(A)))$.

Problem 6

If A is an $m \times n$ matrix then $n = \text{RANK}(A) + \text{DIM}(\text{KER}(A))$.

If A is an $n \times n$ matrix then we say it is full rank iff $\text{RANK}(A) = n$ (or equivalently, iff $\text{DIM}(\text{KER}(A)) = 0$).

A related notion to matrix rank is the rank of an arbitrary subset of vectors, defined as follows.

Definition 2.9 (Rank of a Set of Vectors) *For $V \subseteq \mathbb{R}^n$, we let the rank of V be the maximum size of a linearly independent subset of V, namely*

$$\text{RANK}(V) := \max_{U \subseteq V : U \text{ independent}} |U|.$$

It is not too hard to see that the rank of a matrix is the rank of its columns under the above definition.

Problem 7

For any $m \times n$ matrix A we have $\text{RANK}(\text{A}) = \text{RANK}(\text{ROWS}(\text{A})) = \text{RANK}(\text{COLS}(\text{A}))$.

2.2 Inner Products

One important special case of linear functions that we will study are inner products.

Definition 2.10 (Inner Products) *The inner product of vectors $a, b \in \mathbb{R}^n$ is defined as*

$$\langle a, b \rangle := \sum_i a_i b_i$$

Notice that if we fix vector $a \in \mathbb{R}^n$ and let $f(x) := \langle a, x \rangle$ then for every $x \in \mathbb{R}^n$ we have $f(x) = Ax$ for the $1 \times n$ matrix A whose single row is a. It follows by the above that the inner product is a linear function. In particular, we have the following.

Fact 2.11 *For every $a, x, y \in \mathbb{R}^n$ we have*

$$\langle a, x + y \rangle = \langle a, x \rangle + \langle a, y \rangle$$

and for all $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$ we have

$$\langle a, cx \rangle = c\langle a, x \rangle.$$

We will also often use the fact that the inner product is symmetric.

Fact 2.12 *For any $x, y \in \mathbb{R}^n$ we have*

$$\langle x, y \rangle = \langle y, x \rangle.$$

3 Systems of Linear Equations

One of the most important applications of linear algebra is solving systems of linear equations.

Definition 3.1 (System of Linear Equations) *A system of linear equations consists of $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ and asks whether there exists an $x \in \mathbb{R}^n$ such that*

$$x_1 a_{11} + x_2 a_{12} + x_3 a_{13} + \ldots + x_n a_{1n} = b_1$$
and
$$x_1 a_{21} + x_2 a_{22} + x_3 a_{23} + \ldots + x_n a_{2n} = b_2$$
and
$$x_1 a_{31} + x_2 a_{32} + x_3 a_{33} + \ldots + x_n a_{3n} = b_3$$
and
$$\ldots$$
and
$$x_1 a_{m1} + x_2 a_{m2} + x_3 a_{m3} + \ldots + x_n a_{mn} = b_m$$

or equivalently whether there exists an $x \in \mathbb{R}^n$ such that

$$\langle x, a_i \rangle = b_i \qquad \text{for all } i \in [m]$$

or equivalently whether there exists an $x \in \mathbb{R}^n$ such that

$$Ax = b$$

where A is the matrix whose rows are $a_1, a_2, \ldots, a_m$.

Observe that if $b = 0$ (the all zeros vector) then x is a solution to $Ax = b = 0$ iff it is in $\text{KER}(A)$.

Problem 8

The equivalences in Definition 3.1 are actually equivalent.

A useful characterization of the solution set of a system of linear equations is the following.

Fact 3.2 *Let $K = \{x : Ax = b\}$ be all solutions to $Ax = b$. Then either $K = \emptyset$ or for any $u \in K$ we have $K = u + \text{KER}(A)$.*[1]

The above fact along with the Fundamental Theorem of Linear Maps (Fact 2.4) gives us a nice uniqueness criteria for certain systems of linear equations, as follows.

Problem 9

Given $n \times n$ matrix A, if $Ax = b$ has a solution and $\text{RANK}(A) = n$ then $Ax = b$ has a unique solution.

3.1 Gaussian Elimination

Gaussian elimination is an algorithm which you likely learned in your linear algebra class to solve systems of linear equations.

[1] Here $u + \text{KER}(A) := \{u + v : v \in \text{KER}(A)\}$.

Fact 3.3 *Given an $m \times n$ matrix A and $b \in \mathbb{R}^m$, one can (by Gaussian elimination) in poly(n, m) output an $x \in \mathbb{R}^n$ such that $Ax = b$ or correctly decide that there is no such x.*

If you remember how Gaussian elimination works and you assume that you can add or multiply arbitrarily-large numbers in constant time then the above is quite straightforward to show.[2]

Gaussian elimination is actually a pretty flexible algorithm. For instance, given n vectors $V \subseteq \mathbb{R}^m$, one can use Gaussian elimination to solve a certain set of linear equations to decide if V is independent or not.

Problem 10

Given n vectors $V \subseteq \mathbb{R}^m$, one can in poly(n, m) time decide if V is independent.

[2]There is a slightly annoying issue here, however, because as you run Gaussian elimination the numbers in your matrix can get larger and larger and so if you don't assume that you can add or multiply arbitrarily-large numbers in constant time it's not entirely clear how to actually implement Gaussian elimination in poly(n, m) time. Nonetheless, there are tricks to get around this issue.

Review: Basics of Probability

An Algorithmist's Toolkit (CSCI 2952T)

D Ellis Herskowitz
Brown University

The goal of this is to quickly remind you of the basics of probability, much of which we will later use in class. I'll assume that you've seen most of this before and won't provide proofs. If you're curious about how to prove something, though, feel free to ask!

For each of the problems you are asked to solve, you are free to assume any of the preceding facts in the document (other than the fact itself).

1 Discrete Probability Spaces

The basic object of study of (discrete) probability is discrete probability spaces, as defined below.[1]

Definition 1.1 (Discrete Probability Space) *A discrete probability space consists of*

1. *A (countable) set S called the sample space.*
2. *A probability function $P : S \to [0, 1]$ satisfying $\sum_{o \in S} P(o) = 1$.*

A single element $o \in S$ is referred to as an "outcome" of the sample space.

As an example, if we consider rolling a six-sided die, then a reasonable sample space would be $S = \{1, 2, 3, 4, 5, 6\}$ and if the die is fair then $P(o) = 1/6$ for every $o \in S$.

1.1 Events

Events allow us to formalize the idea of *something* happening in our probability space.

Definition 1.2 (Event) *An event is a subset $E \subseteq S$.*

We can extend the definition of probability from an outcome to an event as follows.

Definition 1.3 (Event Probability) *The probability of event $E \subseteq S$ is*

$$P(E) := \sum_{o \in E} P(o).$$

Given two events E_1 and E_2, the union of these events naturally corresponds to an outcome of E_1 or E_2 occurring and the intersection corresponds to both E_1 and E_2 occurring. For this reason, we will often write "E_1 or E_2" or "E_1 and E_2" to stand in for the events $E_1 \cup E_2$ and $E_1 \cap E_2$ respectively.

The union bound gives a quick and easy upper bound on the "or" of two events.

[1] Recall that a set is countable if it is finite or has cardinality equal to that of the natural numbers.

Theorem 1.4 (Two Event Union Bound) *Given any events E_1, E_2 we have*

$$P(E_1 \cup E_2) \leq P(E_1) + P(E_2).$$

Problem 1

Prove the two event union bound.

More generally, it upper bounds the probability of a collection of events by the sum of their probabilities.

Theorem 1.5 (Fully General Union Bound) *Given any events $E_1, E_2, \ldots, E_k$ we have*

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) \leq \sum_{i=1}^{k} P(E_i).$$

Problem 2

Prove the fully general union bound using the two event union bound.

1.2 Conditioning and Independence

Conditioning on an event allows us to update our probabilities under the assumption that the event happened. In particular, we can define a new probability space assuming event B. The probabilities of this space are given as below.

Definition 1.6 (Conditional Probability of an Outcome) *Given discrete probability space (S, P) and event $B \subseteq S$, we define the conditional probability of outcome o as*

$$P_B(o) := \begin{cases} \frac{P(o)}{P(B)}. & \text{if } o \in B \\ 0 & \text{if } o \notin B. \end{cases}$$

We will refer to (S, P_B) as the result of conditioning on B—as below, this result is itself just a new discrete probability space.

Problem 3

Given any discrete probability space (S, P) and any event $B \subseteq S$, we have that (S, P_B) is a discrete probability space.

We can extend the notion of conditional probabilities from outcomes to events as below.

Definition 1.7 (Conditional Probability of an Event) *Given discrete probability space (S, P) and event $B \subseteq S$, we define the conditional probability of event A as*

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}.$$

It is not too hard to see that the probability of A conditioned on B is just the probability of A in the new probability space (S, P_B).

Problem 4

Prove that for any events A and B we have $P(A \mid B) = P_B(A)$.

Informally, two events are independent if knowing one happened does not change the probability of the other.

Definition 1.8 (Independent Events) *Events A and B are independent iff*

$$P(A \mid B) = P(A)$$

or $P(B) = 0$.

Another common characterization of independent events is as below.

Problem 5

A and B are independent iff $P(A \cap B) = P(A) \cdot P(B)$.

2 Random Variables

Random variables allow us to associate numbers with outcomes.

Definition 2.1 (Random Variable) *A random variable is a function $X : S \to \mathbb{R}$.*

A very common event is one built from a random variable and a predicate as below.[2]

Definition 2.2 (Events from Predicates) *Given a predicate p, we let $p(X)$ be the event*

$$\{o \in S : p(X(o))\}.$$

As an example, "$X = i$" is used to stand for the event $\{o \in S : X(o) = i\}$ and "$X \leq i$" is used to stand for the event $\{o \in S : X(o) \leq i\}$.

2.1 Expectation

The expectation of a random variable is its (weighted) average.

Definition 2.3 (Expectation) *The expectation of random variable X is defined as*

$$\mathbb{E}[X] := \sum_{o \in S} X(o) \cdot P(o).$$

A useful alternate definition of the expectation is as below.

Problem 6

Given any random variable X in a discrete probability space, we have $\mathbb{E}[X] = \sum_i i \cdot P(X = i)$.

Linearity of expectation is one of the most useful algebraic properties of expectation.

[2] Recall that a predicate is a function which takes in values and outputs a proposition—namely, a statement that is either true or false.

Theorem 2.4 (Linearity of Expectation) *For any random variables X and Y and reals $a, b \in \mathbb{R}$ we have*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

In general, expectation does not distribute over multiplication as it does over addition—however, it does if the random variables are independent.

Theorem 2.5 *For any independent random variables X and Y we have*

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Just as conditioning on an event gives us new probabilities for outcomes and events, conditioning on an event gives us new expectations as below. In fact, these expectations are just expectations in the probability space resulting from conditioning.

Definition 2.6 (Conditional Expectation) *The expectation of random variable X conditioned on event B is*

$$\mathbb{E}[X \mid B] := \sum_{o \in B} X(o) \cdot P_B(o).$$

It is not too hard to see that, just as there is an alternate characterization of expectation, there is an alternate characterization of conditional expectation

Theorem 2.7 *Given any random variable X and event B we have*

$$\mathbb{E}[X \mid B] := \sum_{i} i \cdot P(X = i \mid B).$$

2.2 Variance and Covariance

The variance tells us how far, on average, a random variable is from its expectation.

Definition 2.8 (Variance) *The variance of random variable X is*

$$Var(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

A useful alternate characterization of variance is as below.

Theorem 2.9 *For any random variable X we have*

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Problem 7

Show Theorem 2.9.

The covariance tells us how far, on average, two random variables are from each other.

Definition 2.10 (Covariance) *The covariance of two random variables X and Y is*

$$Cov(X, Y) := \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

Variance is not linear like expectation. However, one can pull out reals up to a squaring.

Theorem 2.11 *For any random variables X and real $c \in \mathbb{R}$ we have*

$$Var(cX) = c^2\ Var(X).$$

Likewise, it does distribute across addition up to a covariance term.

Theorem 2.12 *For any random variables X and Y we have*

$$Var(X+Y) = Var(X) + Var(Y) + 2\ Cov(X,Y).$$

Furthermore, if the random variables are independent then covariance is 0 and so it does distribute across addition in this case.

Theorem 2.13 *For any independent random variables X and Y we have*

$$Cov(X,Y) = 0.$$

Problem 8

Show Theorem 2.13.

Theorem 2.14 *For any independent random variables X and Y we have*

$$Var(X+Y) = Var(X) + Var(Y).$$

2.3 PMFs, CDFs and Distributions

The probability mass function (PMF) of a discrete random variable gives the probability that it takes on a particular value.

Definition 2.15 (Probability Mass Function (PMF)) *Given a discrete random variable X, the PMF of X is the function P_X which on input $i \in \mathbb{R}$ gives*

$$P_X(i) = P(X = i).$$

The cumulative distribution function (CDF) gives the probability a discrete random variable takes on at most some value.

Definition 2.16 (Cumulative Distribution Function (CDF)) *Given random variable X, the cumulative distribution function (CDF) of X is the function C_X which on $i \in \mathbb{R}$ gives*

$$C_X(i) := \sum_{a \leq i} P_X(a).$$

The word "distribution" is often used to either describe the PMF or the CDF of a random variable.

It is worth noting that the support of a random variable and its PMF induce a new discrete probability space, as below.

Definition 2.17 (Random Variable Support) *Given discrete probability space (S, P) and a random variable X, the support of X is all values the random variable takes on, namely*

$$supp(X) := \{X(o) : o \in S\}.$$

The following gives the new discrete probability space.

Theorem 2.18 *Given any discrete probability space (S, P) and random variable X, it holds that $(supp(X), P_X)$ is a discrete probability space.*

3 Continuous RVs

For the most part, we will be working with discrete probability spaces. However, there are a few places where we will use continuous probability spaces.

In rough terms, a continuous probability space is one with a sample space that is uncountable—that is, cardinality on the order of $|\mathbb{R}|$. These continuous spaces replace the probability function P with a "probability measure". Furthermore, unlike in the discrete case where we talk about the probability of outcomes and events, here we generally only talk about the probability of an event as specified by the probability measure. Furthermore, the valid events of a continuous spaces are not arbitrary subsets of 2^Ω, but rather a special collection of subsets of 2^Ω which form what is called a σ-algebra.

Random variables likewise get extended. The distribution of random variable X, namely P_X, has as its analogue the probability density function ϕ_X. Likewise, the CDF and expectation generalize but become integrals rather than sums. Like in the discrete case where the PMF sums to 1, the PDF must always integrate to 1.

We won't need the details of this formalism, but the following table compares the two cases.

Discrete Probability Space	**Continuous Probability Space**
Sample space Ω is countable	Sample space Ω is uncountable
Probability function P	Probability Measure P
An event is any subset of 2^Ω	An event is an element of a σ-algebra $\subseteq 2^\Omega$
PMF P_X	PDF ϕ_X
CDF $C_X(i) = \sum_{a \le i} P_X(a)$	CDF $C_X(i) = \int_{-\infty}^{i} \phi_X(a)\, da$
$\mathbb{E}[X] = \sum_i i \cdot P(X = i)$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} i \cdot \phi(i)\ \ di$
$\sum_i P_X(i) = 1$	$\int_{-\infty}^{\infty} \phi(i)\, di = 1$

3.1 Gaussians

The main continuous random variable we will make use of is the Gaussian.

Definition 3.1 (Gaussian PDF) *A continuous random variable X is said to be Gaussian if its PDF on $i \in \mathbb{R}$ is*

$$\phi_X(i) = \frac{1}{\sqrt{2\pi}} e^{-i^2/2}.$$

It is a surprisingly interesting argument to show that the PDF of a Gaussian indeed integrates to 1.

Theorem 3.2 (Gaussian PDF is Actually a PDF) *Letting ϕ be the Gaussian PDF as in Definition 3.1, we have*

$$\int_{-\infty}^{\infty} \phi(x)\,dx = 1.$$

Since ϕ_X is symmetric around the y-axis, it is easy to check that if X is Gaussian then $\mathbb{E}[X] = 0$. Likewise, with a bit more work one can show that $\mathrm{Var}(X) = 1$ for a Gaussian random variable. As such, if X is a random variable with a Gaussian PDF, then we write $X \sim N(0,1)$. Such an X is referred to as a standard Gaussian.

A non-standard Gaussian is as follows.

Definition 3.3 (Non-Standard Gaussian) *Y is a non-standard Gaussian if it is of the form $\mu + \sigma X$ where $X \sim N(0,1)$ is a standard Gaussian.*

It is easy to check that such a Gaussian has expectation and variance μ and σ^2 respectively, hence for such a Gaussian we write $Y \sim N(\mu, \sigma^2)$.

We will make use of 2 key properties of Gaussians. The first of these states that the direction that a vector whose entries are iid Gaussians is uniformly random. Equivalently, any two vectors with the same norm are equally likely when coordinates are drawn as iid Gaussians (you should convince yourself this is equivalent).[3]

Theorem 3.4 (Rotational Symmetry of Gaussians) *Suppose that $u = (u_1, u_2, \ldots, u_k)$ is a random vector in $\mathbb{R}^k$ where each $u_i \sim N(0,1)$ independently. Then for any $x, y \in \mathbb{R}^k$ with $\|x\| = \|y\|$ we have*

$$P(u = x) = P(u = y).$$

Second, if we sum together two scaled Gaussians, the result is a Gaussian.

Theorem 3.5 (Sum of Gaussians is Gaussian) *If $X \sim N(0,1)$ and $Y \sim N(0,1)$ and X and Y are independent then, for any reals $a, b \in \mathbb{R}$, we have $aX + bY \sim N(0, a^2 + b^2)$*

It's a fun (but tricky) exercise to show Theorem 3.5 using Theorem 3.4.

[3] This is actually somewhat incorrect as stated. In particular, for a continuous random vector the probability of taking any exact value is zero, so expressions like $P(u = x)$ should be interpreted informally.