CCS Discrete II | Professor: Padraic Bartlett

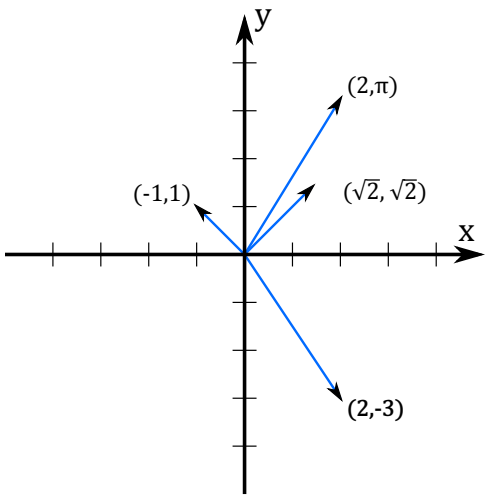
Lecture 2: Vector Spaces, Metric Spaces

Week 2 | *UCSB 2015*

1 Vector Spaces, Informally

The two vector spaces[1] you're probably the most used to working with, from either your previous linear algebra classes or even your earliest geometry/precalc classes, are the spaces $\mathbb{R}^2$ and $\mathbb{R}^3$. We briefly review how these two vector spaces work here:

Definition. The **vector space** $\mathbb{R}^2$ consists of the collection of all pairs (a, b), where a, b are allowed to be any pair of real numbers. For example, $(2, -3), (2, \pi), (-1, 1)$, and $(\sqrt{2}, \sqrt{2})$ are all examples of vectors in $\mathbb{R}^2$. We typically visualize these vectors as arrows in the xy-plane, with the tail of the arrow starting at the origin[2] and the tip of the arrow drawn at the point in the plane with xy-coördinates given by the vector. We draw four such vectors here:

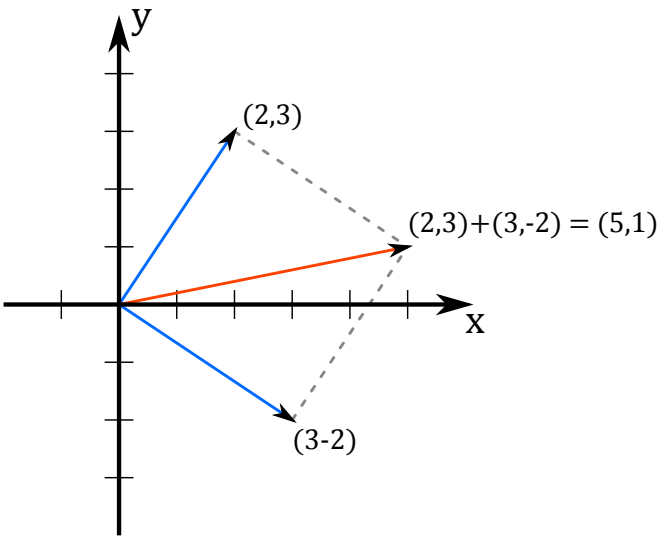


Given a pair of vectors in $\mathbb{R}^2$, we can **add** them together. We do this component-wise, i.e. if we have two vectors (a, b) and (c, d), their sum is the vector $(a+c, b+d)$. For example, the sum of the vectors $(3, -2)$ and $(2, 3)$ is the vector $(5, 1)$.

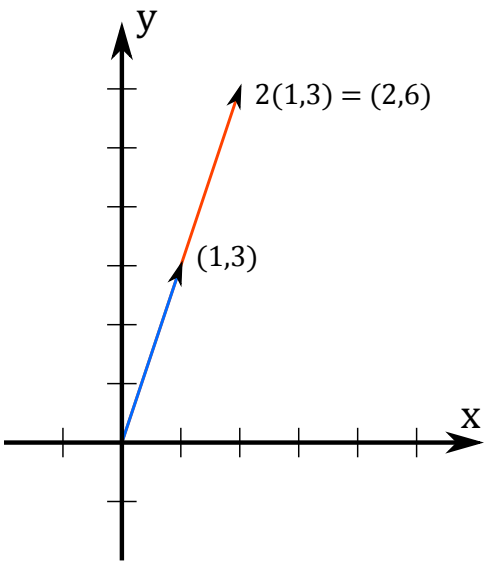
You can visualize this by taking the arrow corresponding to the first vector that we add, and "translating" this arrow over to the start of the second vector; if you travel along the first vector and then continue along this second translated vector, you arrive at some point in the plane. The arrow connecting the origin to this point is the vector given by the sum of these two vectors! If this seems hard to understand, the diagram below may help some:

[1] This section is duplicated from last quarter's Introduction to Proof notes; so, if you were OK with vector spaces back then, this should be OK here as well! Feel free to skip to the next section, which discusses metric spaces.

[2] The origin is the point $(0, 0)$ in the plane.

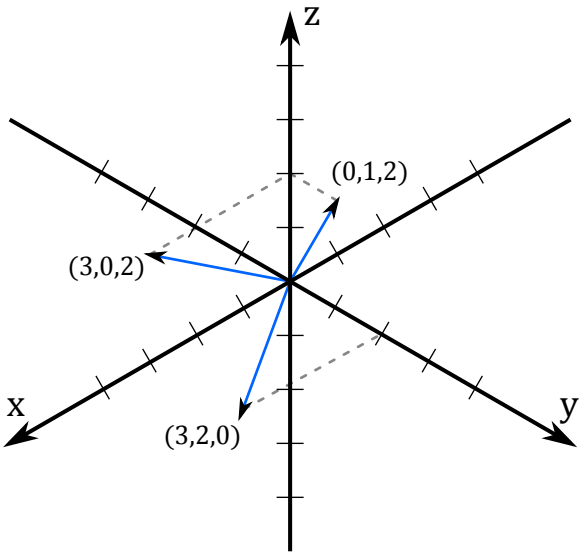


We can also **scale** a vector in $\mathbb{R}^2$ by any real number a. Intuitively, this corresponds to the concept of "stretching:" the vector (x, y) scaled by a, denoted $a(x, y)$, is the quantity (ax, ay). For example, $2(1, 3) = (2, 6)$, and is essentially what happens if we "double" the vector $(1, 3)$. We illustrate this below:

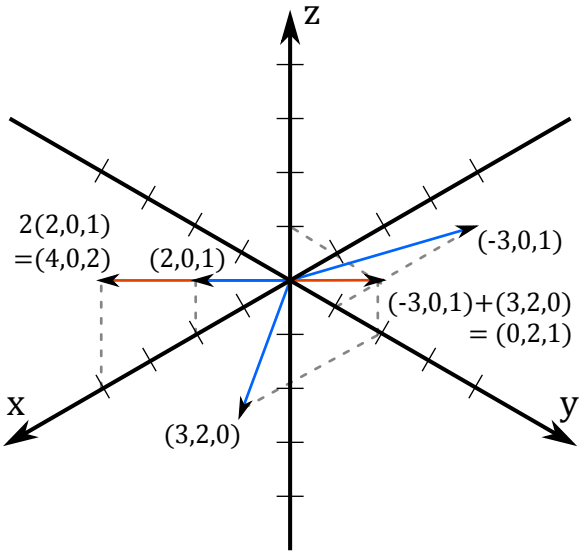


We can define $\mathbb{R}^3$ in a similar fashion:

Definition. The **vector space** $\mathbb{R}^3$ consists of the collection of all pairs (a, b, c), where a, b, c are allowed to be any triple of real numbers. For example, $(0, 1, 2), (3, 0, 2)$, and $(3, 2, 0)$ are all examples of vectors in $\mathbb{R}^3$. We typically visualize these vectors as arrows in three-dimensional xyz-space, with the tail of the arrow starting at the origin and the tip of the arrow drawn at the point in the plane with xyz-coördinates given by the vector. We draw three such vectors here:



Again, given a pair of vectors in $\mathbb{R}^3$, we can **add** them together. We do this component-wise, i.e. if we have two vectors (a, b, c) and (d, e, f), their sum is the vector $(a+d, b+e, c+f)$. For example, the sum of the vectors $(3, -2, 0)$ and $(2, 1, 2)$ is the vector $(5, -1, 2)$. We can also **scale** a vector in $\mathbb{R}^3$ by any real number a: the vector (x, y, z) scaled by a, denoted $a(x, y, z)$, is the quantity (ax, ay, az). These operations can be visualized in a similar fashion to the pictures we drew for $\mathbb{R}^2$:



You can generalize this discussion to $\mathbb{R}^n$, the vector space made out of n-tuples of real numbers: i.e. elements of $\mathbb{R}^4$ would be things like $(\pi, 2, 2, 1)$ or $(-1, 2, 1, -1)$. In fact, you can generalize this entire process to any arbitrary field F:

Definition. Take any set S. The **n-fold** Cartesian product of S with itself is the collection of all ordered n-tuples of elements of S: that is,

$$S^n = \{(s_1, s_2, \ldots s_n) \mid s_1, s_2, \ldots s_n \in S\}$$

Suppose specifically that S is actually some field F (examples of fields we studied last quarter: $\mathbb{R}, \mathbb{C}, \mathbb{Q}, \mathbb{Z}/p\mathbb{Z}$ whenever p is a prime, $\mathbb{F}_q$ for any prime power q via our equivalence-relation constructions.) We can define the operation of vector addition $+ : F^n \times F^n \to F^n$ on this set as follows: for any $\vec{f} = (f_1, \ldots f_n), \vec{g} = (g_1, \ldots g_n) \in F^n$, we can form

$$(f_1, f_2, \ldots f_n) + (g_1, g_2, \ldots g_n) := (f_1 + g_1, f_2 + g_2 + \ldots f_n + g_n),$$

where the addition done on the right-hand side above is done using F's addition operation. Similarly We can also define the operation of scalar multiplication $\cdot : F \times F^n \to F^n$ as follows: for any $\vec{f} \in F^n, a \in F$, we can form the vector

$$a(f_1, f_2, \ldots f_n) = (a \cdot f_1, a \cdot f_2, \ldots a \cdot f_n),$$

where again the multiplication done on the right-hand side is done using F's multiplication operation.

2 Vector Spaces, Formally

In general, there are many other kinds of vector spaces — essentially, anything with the two operations "addition" and "scaling" is a vector space, provided that those operations are well-behaved in certain specific ways. Much like we did with $\mathbb{R}$ and the field axioms, we can generate a list of "properties" for a vector space that seem like characteristics that will insure this "well-behaved" nature. We list a collection of such properties and use them to define a vector space here:

Definition. A **vector space** V over a field F is a set V along with the two operations addition and scalar multiplication, such that the following properties hold:

- **Closure**$(+)$: $\forall \vec{v}, \vec{w} \in V$, we have $v + w \in V$.
- **Identity**$(+)$: $\exists \vec{0} \in V$ such that $\forall \vec{v} \in V$, $\vec{0} + \vec{v} = \vec{v}$.
- **Commutativity**$(+)$: $\forall \vec{v}, \vec{w} \in V, \vec{v} + \vec{w} = \vec{w} + \vec{v}$.
- **Associativity**$(+)$: $\forall \vec{u}, \vec{v}, \vec{w} \in V, (\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
- **Inverses**$(+)$: $\forall \vec{v} \in V, \exists$ some $-\vec{v} \in V$ such that $\vec{v} + (-\vec{v}) = 0$.
- **Closure**$(\cdot)$: $\forall a \in F, \vec{v} \in V$, we have $a\vec{v} \in V$.
- **Identity**$(\cdot)$: $\forall \vec{v} \in V$, we have $1\vec{v} = \vec{v}$.
- **Compatibility**$(\cdot)$: $\forall a, b \in F$, we have $a(b\vec{v}) = (a \cdot b)\vec{v}$.
- **Distributivity**$(+, \cdot)$: $\forall a \in F, \vec{v}, \vec{w} \in V, a(\vec{v} + \vec{w}) = a\vec{v} + a\vec{w}$.

As with fields, there are certainly properties that $\mathbb{R}^n$ satisfies that are not listed above. For example, consider the following property:

- **New property?**$(+)$: The additive identity, $\vec{0}$, is unique in any vector space. In other words, there cannot be two distinct vectors that are both the additive identity for a given vector space.

Just like before, this property turns out to be redundant: in other words, this property is implied by the definition of a vector space! We prove this here:

Claim. In any vector space, the additive identity is unique.

Proof. Take any two elements $\vec{0}, \vec{0'}$ that are both additive identities. Then, by definition, we know that because $\vec{0}$ is an additive identity, we have

$$\vec{0'} = \vec{0} + \vec{0'}.$$

Similarly, because $\vec{0'}$ is an additive identity, we have

$$\vec{0} = \vec{0'} + \vec{0}.$$

If we use commutativity to switch the $\vec{0}$ and $\vec{0'}$, we can combine these two equalities to get that

$$\vec{0} = \vec{0'} + \vec{0} = \vec{0} + \vec{0'} = \vec{0'}.$$

Therefore, we have shown that $\vec{0}$ and $\vec{0'}$ are equal. In other words, we've shown that all of the elements that are additive identities are all equal: i.e. that they're all the same element! Therefore, this additive identity element is **unique**: there is no other element that is somehow an additive identity that is different from $\vec{0}$. □

As we did with fields, there are a number of other properties that $\mathbb{R}^n$ possesses that you can prove that any vector space must have: in your textbook, there are proofs that every vector has a unique additive inverse, that $0\vec{v}$ is always $\vec{0}$, that $-1\vec{v} = -\vec{v}$, and other such things.

Instead of focusing on more of these proofs, we shift our attention instead to actually describing some vector spaces!

A few of these are relatively simple to come up with:

- $\mathbb{R}^n$, the example we used to come up with these properties, is a vector space over the field $\mathbb{R}$.
- $\mathbb{C}^n$ is similar. Specifically: $\mathbb{C}^n$ is the set of all n-tuples of complex numbers: i.e.

  $$\mathbb{C}^n = \{(z_1, \dots z_n) | z_1, \dots z_n \in \mathbb{C}\}.$$

  Just like with $\mathbb{R}^n$, we can add these vectors together and scale them by arbitrary complex numbers, while satisfying all of the vector space properties. We leave the details for the reader to check, but this is a vector space over the complex numbers $\mathbb{C}$.
- Similarly, $\mathbb{Q}^n$, the set of all n-tuples of rational numbers

  $$\mathbb{Q}^n = \{(q_1, \dots q_n) | q_1, \dots q_n \in \mathbb{Q}\},$$

  is a vector space over the field $\mathbb{Q}$.

- In general, given any field F, the set F^n along with the vector addition and scalar multiplication operations defined earlier, is a vector space!

  This is not hard to check:

  - **Closure(+)**: Immediate. Because F is a field and is closed under addition, the pairwise sums performed in vector addition must create another vector.
  - **Identity(+)**: Because F is a field, it has an additive identity, 0. The vector $\vec{0} = (0, 0, \ldots 0)$ is consequently the additive identity for our vector space, as pairwise adding this vector to any other vector does not change any of the other vector's coördinates.
  - **Commutativity(+)**: Again, this is a consequence of F being a vector space. Because addition is commutative in F, the pairwise addition in our vector space is commutative.
  - **Associativity(+)**: Once more, this is a consequence of F being a vector space. Because addition is associative in F, the pairwise addition in our vector space is associative.
  - **Inverses(+)**: Take any $\vec{f} = (f_1, \ldots f_n) \in F^n$. Because F is a field, we know that $(-f_1, \ldots - f_n)$ is a vector in F^n as well. Furthermore, the pairwise addition of these two vectors clearly yields the additive identity $\vec{0}$; therefore, our vector space has inverses.
  - **Closure($\cdot$)**: This is a consequence of F being closed under multiplication.
  - **Identity($\cdot$)**: Because F is a field, it has a multiplicative identity 1. This 1, when used to scale a vector, does not change that vector at any coördinate because of this multiplicative identity property; therefore 1 is also the scalar multiplicative identity for our vector space.
  - **Compatibility($\cdot$)**: This is an immediate consequence from F's multiplication being associative, as for any $a, b \in F$, we have

    $$\begin{aligned} a(b(f_1 \ldots f_n) =& a(b \cdot f_1, \ldots b \cdot f_n) = (a \cdot (b \cdot f_1), \ldots a \cdot (b \cdot f_n)) \\ =& (a \cdot b) \cdot f_1, \ldots (a \cdot b) \cdot f_n) = (a \cdot b)(f_1, \ldots f_n). \end{aligned}$$

  - **Distributivity($+, \cdot$)**: This is a consequence of F being a vector space. Because multiplication and addition are distributive in F, their combination in our vector space is distributive as well.

- A specific consequence of the above result is that something like $(\mathbb{Z}/5\mathbb{Z})^n$ is a vector space. This is a somewhat strange-looking beast: it's a vector space over a finite-sized field! In particular, it's a vector space with only finitely many elements, which is weird.

  To understand this better, we look at some examples. Consider $(\mathbb{Z}/5\mathbb{Z})^2$. This is the vector space consisting of elements of the form

  $$(a, b),$$

where $a, b \in \{0, 1, 2, 3, 4\}$. We add and scale elements in this vector space using mod-5 modular arithmetic: for example,

$$(2, 3) + (4, 4) = (1, 2),$$

because $2 + 4 \equiv 1 \mod 5$ and $3 + 4 \equiv 2 \mod 5$. Similarly,

$$2(3, 1) = (1, 2),$$

because $2 \cdot 3 \equiv 1 \mod 5$ and $2 \cdot 1 \equiv 2 \mod 5$.

Perhaps surprisingly, these odd-looking vector spaces are some of the most-commonly used spaces in the theoretical computer science/cryptographic settings. In particular, they come up very often in the field of elliptic curve cryptography, as you may remember from last quarter!

There are some odder examples of vector spaces:

- Polynomials! Specifically, let $\mathbb{R}[x]$ denote the collection of all finite-degree polynomials in one variable x with real-valued coefficients. In other words,

  $$\mathbb{R}[x] = \{a_0 + a_1 x + \ldots a_n x^n | a_0, \ldots a_n \in \mathbb{R}, n \in \mathbb{N}\}.$$

  Verifying that this is a vector space is not very difficult:

  - **Closure**$(+)$: Adding two polynomials together clearly gives us another polynomial.
  - **Identity**$(+)$: Adding 0 to any polynomial doesn't change it, and 0 is a polynomial itself (simply pick $a_0 = 0$ and $n = 0$.)
  - **Commutativity**$(+)$: We can add polynomials in any order that we want, and we'll always get the same answer. (This is because addition in $\mathbb{R}$ is commutative, and we just add polynomials by grouping common powers of x and adding their real-valued coefficients together!)
  - **Associativity**$(+)$: Holds for the precise same reason that commutativity holds.
  - **Inverses**$(+)$: Given any polynomial $a_0 + \ldots a_n x^n$, the polynomial $-a_0 + \ldots - a_n x^n$ is its additive inverse, as summing these two polynomials gives us 0.
  - **Closure**$(\cdot)$: Multiplying a polynomial by a real number clearly gives us another polynomial.
  - **Identity**$(\cdot)$: Multiplying a polynomial by 1 clearly gives us the same polynomial back.
  - **Distributivity**$(+, \cdot)$: Holds for the precise same reason that commutativity holds.

- Matrices! Specifically, let $M_{\mathbb{R}}(n, n)$ denote the set of $n \times n$ matrices with real-valued entries. For example

  $$M_{\mathbb{R}}(3, 3) = \left\{ \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \;\middle|\; a, b, c, d, e, f, g, h, i \in \mathbb{R} \right\}.$$

If we define matrix addition as simply entrywise addition: i.e.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \dots & a_{nn}+b_{nn} \end{bmatrix},$$

and scalar multiplcation as simply entrywise multiplication, i.e.

$$c\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \dots & ca_{nn} \end{bmatrix},$$

then this is a vector space! Specifically, it's a vector space for precisely the same reasons that $\mathbb{R}^n$ is a vector space: if you just think of a $n \times n$ matrix as a very oddly-written vector in $\mathbb{R}^{n^2}$, then every argument for why $\mathbb{R}^{n^2}$ is a vector space carries over to $M_{\mathbb{R}}(n, n)$.

It might seem odd to think of matrices as a vector space, but if you go further in physics or pure mathematics, this is an incredibly useful and common construction. We leave the details of checking this is a vector space to the reader, but the proof works just like everything else we've done thus far!

In our last lecture, the notation of "vector spaces" is not the only object we used when studying/creating error-correcting codes. We also needed to use the concept of **distance** when studying our codes; however, when we did this, we used a very strange notion of distance (the Hamming distance) instead of the normal notion of distance that we have from $\mathbb{R}^n$.

This raises a natural question: what **other** notions of distance exist? Are there any properties common to all notions of distance? We study these questions here.

3 Metric Spaces

Definition. Take any set S. A **metric** on S is any function $d : S \times S \to \mathbb{R}$ such that the following four properties hold:

1. **Identity of indiscernible values**[3]: For any $x \in S$, we have $d(x, x) = 0$.
2. **Positivity**: For any two distinct $x, y \in S$, we have $d(x, y) > 0$.
3. **Symmetry**: For any $x, y \in S$, we have $d(x, y) = d(y, x)$.
4. **Triangle inequality**: For any $x, y, z \in S$, we have $d(x, z) \leq d(x, y) + d(y, z)$.

[3]No one uses this term.

The **Euclidean distance** is probably the first metric you encountered in your life. To be rigorous, we define the **Euclidean distance** on $\mathbb{R}^n$ between any two vectors $\vec{x}, \vec{y}$ by

$$d(\vec{x}, \vec{y}) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}.$$

We start by rigorously showing that this is a metric:

Claim. The Euclidean distance is a metric on $\mathbb{R}^n$.

Proof. We check properties one-by-one.

Identity of indiscernible values: Take any $\vec{x} \in \mathbb{R}^n$. Then, by definition, we have

$$d(\vec{x}, \vec{x}) = \sqrt{(x_1 - x_1)^2 + \ldots + (x_n - x_n)^2} = \sqrt{0 + \ldots + 0} = 0,$$

as claimed.

Positivity: Take any two distinct $\vec{x}, \vec{y} \in \mathbb{R}^n$. Because $\vec{x} \neq \vec{y}$ by assumption, there must be some coördinate k such that $x_k \neq y_k$. Consequently, because $(x_i - y_i)^2$ is a square and thus no smaller than 0 for any i, we have

$$\begin{aligned}\sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2} &\geq \sqrt{0 + \ldots + 0 + (x_k - y_k)^2 + 0 + \ldots + 0} \\ &= |x_k - y_k| \\ &> 0,\end{aligned}$$

which proves our claim.

Symmetry: Take any two $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then, we have that

$$\begin{aligned}d(\vec{x}, \vec{y}) &= \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2} \\ &= \sqrt{(y_1 - x_1)^2 + \ldots + (y_n - x_n)^2} \\ &= d(\vec{y}, \vec{x}).\end{aligned}$$

Triangle inequality. This proof turns out to be a bit of a pain, which is why we skipped it in class: but if you're curious, read on!

We start by establishing the Cauchy-Schwarz inequality, which (for $\mathbb{R}^n$) states that for any two vectors $\vec{x}, \vec{y}$ we have

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq \sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}.$$

This is a little messy, but not too hard to show. Because the LHS and RHS above are both positive, we can square both sides above and get an equivalent expression:

$$\left(\sum_{i=1}^{n} x_i y_i \right)^2 \leq \left(\sum_{i=1}^{n} x_i^2 \right) \cdot \left(\sum_{i=1}^{n} y_i^2 \right).$$

To prove that this expression works, consider the following algebraic expression, chosen cleverly to relate to both the RHS and LHS:

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(x_iy_j - x_jy_i)^2\right).$$

We can expand this as follows:

$$\begin{aligned}
\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(x_iy_j - x_jy_i)^2\right) &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n}x_i^2y_j^2 + x_j^2y_i^2 - 2x_ix_jy_iy_j\right) \\
&= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^2y_j^2\right) + \left(\sum_{i=1}^{n}\sum_{j=1}^{n}x_j^2y_i^2\right) - 2\left(\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_jy_iy_j\right) \\
&= 2\left(\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^2y_j^2\right) - 2\left(\sum_{i=1}^{n}x_iy_i\cdot\left(\sum_{j=1}^{n}x_jy_j\right)\right) \\
&= 2\left(\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}y_j^2\right) - 2\left(\sum_{i=1}^{n}x_iy_i\cdot\left(\sum_{j=1}^{n}x_jy_j\right)\right) \\
&= 2\left(\sum_{i=1}^{n}x_i^2\right)\left(\sum_{i=1}^{n}y_i^2\right) - 2\left(\sum_{i=1}^{n}x_iy_i\right)\left(\sum_{j=1}^{n}x_jy_j\right) \\
&= 2\left(\sum_{i=1}^{n}x_i^2\right)\cdot\left(\sum_{i=1}^{n}y_i^2\right) - 2\left(\sum_{i=1}^{n}x_iy_i\right)^2.
\end{aligned}$$

We know that the LHS here is a positive number, as it is the sum of many squared real numbers. Consequently, the RHS is also positive; in other words, we must have

$$\left(\sum_{i=1}^{n}x_iy_i\right)^2 \leq \left(\sum_{i=1}^{n}x_i^2\right)\cdot\left(\sum_{i=1}^{n}y_i^2\right).$$

Taking square roots gives us

$$\left|\sum_{i=1}^{n}x_iy_i\right| \leq \sqrt{\sum_{i=1}^{n}x_i^2}\cdot\sqrt{\sum_{i=1}^{n}y_i^2}.$$

as claimed.

Notationally, people usually simplify this by letting $||\vec{x}|| = \sqrt{x_1^2 + \ldots x_n^2}$, i.e. $||\vec{x}||$ denotes the distance from $\vec{0}$ to the point $\vec{x}$. If you use this above, our inequality simplifies to

$$\left|\sum_{i=1}^{n}x_iy_i\right| \leq ||\vec{x}||\cdot||\vec{y}||.$$

We can use the Cauchy-Schwarz inequality to prove the following useful lemma:

Lemma. Take any two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then $||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + \vec{y}||$.

Proof. We just do some algebraic manipulations, and use Cauchy-Schwarz:

$$\begin{aligned}
||\vec{x} + \vec{y}||^2 &= \left(\sqrt{(x_1 + y_1)^2 + \ldots + (x_n + y_n)^2}\right)^2 \\
&= (x_1 + y_1)^2 + \ldots + (x_n + y_n)^2 \\
&= \left(\sum_{i=1}^{n} x_i^2\right) + \left(\sum_{i=1}^{n} y_i^2\right) + 2\left(\sum_{i=1}^{n} x_i y_i\right) \\
&= ||\vec{x}||^2 + ||\vec{y}||^2 + 2\left(\sum_{i=1}^{n} x_i y_i\right) && \leq ||\vec{x}||^2 + ||\vec{y}||^2 + 2||\vec{x}|| \cdot ||\vec{y}|| \\
&= \left(||\vec{x}|| + ||\vec{y}||\right)^2
\end{aligned}$$

Taking square roots proves our claim! □

With this lemma, we can prove the Cauchy-Schwarz inequality. Take any three $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$. We want to prove that $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$.

To do this, simply notice that

$$\begin{aligned}
d(\vec{x}, \vec{z}) &= \left(\sqrt{(x_1 - z_1)^2 + \ldots + (x_n - z_n)^2}\right) \\
&= ||\vec{x} - \vec{z}|| \\
&= ||\vec{x} - \vec{y} + \vec{y} - \vec{z}|| \\
&\leq ||\vec{x} - \vec{y}|| + ||\vec{y} - \vec{z}|| \\
&= d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z}).
\end{aligned}$$

Success! (In your analysis classes you'll probably do this in class. Let me know when you do?) □

This is not the only metric that people use! Another useful metric is the **taxicab metric**, defined below:

Definition. The **taxicab metric** on $\mathbb{R}^n$ is defined as follows: for any $\vec{x}, \vec{y} \in \mathbb{R}^n$, we have

$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

The idea with this metric is that if you're in a large city and need to drive someone from one point to the other, if your city is set up with a grid system, you can only move along one coördinate at a time! So, for example, to go one mile north and one mile east in a city like Chicago, you'd likely have to drive two miles (one north and one east) rather than $\sqrt{2}$ miles northeast, because that diagonal road probably doesn't exist.

This, too, is a metric:

Claim. The taxicab metric is a metric on $\mathbb{R}^n$.

Proof. **Identity of indiscernible values**: Take any $\vec{x} \in \mathbb{R}^n$. Then, by definition, we have

$$d(\vec{x}, \vec{x}) = \sum_{i=1}^{n} |x_i - x_i| = \sum_{i=1}^{n} 0 = 0,$$

as claimed.

Positivity: Take any two distinct $\vec{x}, \vec{y} \in \mathbb{R}^n$. Because $\vec{x} \neq \vec{y}$ by assumption, there must be some coördinate k such that $x_k \neq y_k$. Consequently, because $|x_i - y_i| > 0$ by definition, we have

$$\sum_{i=1}^{n} |x_i - y_i| \geq |x_i - y_i| > 0,$$

which proves our claim.

Symmetry: Take any two $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then, we have that

$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} |y_i - x_i| = d(\vec{y}, \vec{x}),$$

which proves our claim.

Triangle inequality: Take any three $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$. We prove the triangle inequality here by appealing to the Euclidean metric, which is equal [4]to the taxicab metric on $\mathbb{R}^1$. Simply note that

$$\begin{aligned}
d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z}) &= \sum_{i=1}^{n} |x_i - y_i| + \sum_{i=1}^{n} |y_i - z_i| \\
&= \sum_{i=1}^{n} (|x_i - y_i| + |y_i - z_i|) \\
&\geq \sum_{i=1}^{n} |x_i - y_i + y_i - z_i|, \text{ by Euclidean triangle ineq. for each coordinate} \\
&= \sum_{i=1}^{n} |x_i - z_i| \\
&= d(\vec{x}, \vec{z}).
\end{aligned}$$

So we've proven our claim! □

A third useful metric is the **maximum norm**:

Definition. The **maximum norm** on $\mathbb{R}^n$ is defined as follows: for any $\vec{x}, \vec{y} \in \mathbb{R}^n$, we have

$$d(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i|$$

[4] Proof of this claim: when $n = 1$, $d_{\text{Euclidean}}(x, y) = \sqrt{(x-y)^2} = |x - y| = d_{\text{taxicab}}(x, y)$.

This is a metric as well:

Claim. The taxicab metric is a metric on $\mathbb{R}^n$.

Proof. Like before, the positivity/identity of indiscernible values/symmetry properties all fall out from the definition, leaving only the triangle inequality as the interesting case to check.

We do this here. Take any three $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$. Suppose that j is the coördinate at which $\vec{x}, \vec{z}$ have the greatest difference; then we must have $d(\vec{x}, \vec{z}) = |x_j - z_j|$. By again using the one-dimensional Euclidean norm, we can see that this is no greater than $|x_j - y_j| + |y_j - z_j|$; in turn, we can see that this sum is no greater than the maximums $\left(\max_{1 \le i \le n} |x_i - y_i|\right) + \left(\max_{1 \le i \le n} |y_i - z_i|\right)$. But this is just $d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$; so we've proven our claim! □

We give one last example of a metric on $\mathbb{R}^n$ here, called the "post office" metric:

Definition. The **post office metric** on $\mathbb{R}^n$ is defined as follows:

- For any $\vec{x} \in \mathbb{R}^n$, we have $d(\vec{x}, \vec{x}) = 0$.
- For any $\vec{x} \neq \vec{y} \in \mathbb{R}^n$, we have $d(\vec{x}, \vec{y}) = ||\vec{x}|| + ||\vec{y}||$. (If you haven't seen this notation before, $||\vec{x}||$ denotes the quantity $\sqrt{x_1^2 + \ldots x_n^2}$.)

The idea with this metric is that if you want to mail a package from some location $\vec{x}$ to another location $\vec{y}$, you often have to route it through some central packing location (in this case, $\vec{0}$.)

On the homework, you are asked to prove that this is a metric!

All of the metrics above are defined on $\mathbb{R}^n$. This is not always the case; indeed, many metrics are defined on other sets! The **discrete** metric is a useful example of a metric that can be applied to any set:

Definition. Take any set S. The **discrete metric** on S is defined as follows:

- For any $s \in S$, we have $d(s, s) = 0$.
- For any $s \neq t \in S$, we have $d(s, t) = d(t, s) = 1$.

This is a metric, as we show here:

Claim. The discrete metric is a metric on any set S.

Proof. As before, our metric satisfies the positivity/identity of indiscernible values/symmetry properties by definition; so it suffices to check the triangle inequality. Take any three $r, s, t \in S$; we want to show that

$$d(r, t) \le d(r, s) + d(s, t).$$

This is not hard to do via casework. If $r = t$, then the LHS above is 0, and we are done because our function is nonnegative. Alternately, if $r \neq t$, then the LHS is 1. Consider possible values of s; because $r \neq t$, s can be equal to at most one of r, t. Consequently, at least one term on the RHS is 1 as well, and thus by nonnegativity our inequality holds. □

Finally, we should mention the notion of distance that set up this entire discussion: the **Hamming distance**, which we defined on codewords in $(\mathbb{Z}/q\mathbb{Z})^n$! As it turns out, this too is a metric:

Claim. The Hamming distance is a metric on $(\mathbb{Z}/q\mathbb{Z})^n$.

Proof. We first restate the Hamming distance, in case you've forgotten the definition:

Definition. Take any two words $\vec{v} = (v_1, \dots v_n), \vec{w} = (w_1, \dots w_n)$ from $(\mathbb{Z}/q\mathbb{Z})^n$. The **Hamming distance** between $\vec{v}, \vec{w}$, denoted $d(\vec{v}, \vec{w})$, is the number of places at which these two words disagree. To be formal, define $\chi : (\mathbb{Z}/q\mathbb{Z})^2 \to \{0, 1\}$ by $\chi(a, b) = 1$ if and only if $a \neq b$, and 0 otherwise. Then $d(\vec{v}, \vec{w})$ is just the sum $\sum_{i=1}^{n} \chi(v_i, w_i)$.

With this restated, we can once again see that our definition trivially satisfies the positivity/identity of indiscernible values/symmetry properties! So, as before, the only interesting property to verify is the triangle inequality.

Take any three words $\vec{u}, \vec{v}, \vec{w}$ from $(\mathbb{Z}/q\mathbb{Z})^n$. Suppose that $d(\vec{u}, \vec{v})$ is equal to some k; then by definition, we can turn $\vec{u}$ into $\vec{v}$ by changing exactly k of its entries. Similarly, suppose that $d(\vec{v}, \vec{w})$ is equal to some l; then by definition, we can turn $\vec{v}$ into $\vec{w}$ by changing exactly l of its entries.

Consequently, we can turn $\vec{u}$ into $\vec{w}$ by changing at most $k+l$ of the entries in $\vec{u}$. But this means that $\vec{u}, \vec{w}$ must agree on all but at least $k+l$ places if we only need to change at most $k+l$ places to turn $\vec{u}$ into $\vec{w}$! In other words, we must have $d(\vec{u}, \vec{w} \leq k+l = d(\vec{u}, \vec{v}) + d(\vec{v}, \vec{w})$, as claimed. □

4 Using Metrics to Visualize Codes

The main reason we talk about metrics here is because they give us excellent ways to visualize codes in space. To make this concrete, consider the following definition and claim:

Definition. Take any set S with a metric $d : S \times S \to \mathbb{R}$ defined on it. Take any point $x \in S$ and any radius $r \in \mathbb{R}$. We define the **ball** of radius r around the point x with respect to the metric d as the following set:

$$B_r(x) = \{y \in S \mid d(x, y) \leq r\}.$$

Claim. Take any block-length n q-ary code C. The following two properties are equivalent:

1. The collection of balls $\{B_k(\vec{c}) \mid \vec{c} \in C\}$ are all disjoint.
2. $d(C) \geq 2k + 1$.

Proof. As this is an equivalence claim, we must prove both directions.

$(1 \Rightarrow 2)$: Take any k such that the collection of balls

$$\{B_k(\vec{c}) \mid \vec{c} \in C\}$$

are all disjoint. Consider any two codewords $\vec{c_1}, \vec{c_2} \in C$. If $d(\vec{c_1}, \vec{c_2}) \leq 2k$, then the two balls of radius k centered around $\vec{c_1}, \vec{c_2}$ would overlap. To be specific, our balls will overlap at any word $\vec{w}$ formed by taking the first k places at which $\vec{c_1}$ disagrees with $\vec{c_2}$ and changing those entries to $\vec{c_2}$'s entries, as this word is (by construction) distance k from $\vec{c_1}$, and distance $d(\vec{c_1}, \vec{c_2}) - k \leq 2k - k = k$ from $\vec{c_2}$.

So this cannot happen! In other words, any two codewords in our code must be distance at least $2k + 1$ apart; that is, $d(C) \geq 2k + 1$.

$(2 \Rightarrow 1)$: We basically reverse all of the above steps. Take any code C with $d(C) \geq 2k+1$, and any two codewords $\vec{c_1}, \vec{c_2} \in C$. If the two balls $B_k(\vec{c_1}), B_k(\vec{c_2})$ were to intersect, then by the logic above our two codewords would be distance at most $2k$ from each other, a contradiction to our assumption! So these balls must not intersect, as claimed. $\square$

In this sense, finding "good" codes — i.e. codes with $d(C)) \geq 2k+1$ for some set k, with as many elements as possible — is equivalent to the task of packing $(\mathbb{Z}/q\mathbb{Z})^n$ with balls of radius k! This gives us a nice visualization for what "good" codes look like: ways to place balls in space!