Math 1b | TA: Padraic Bartlett

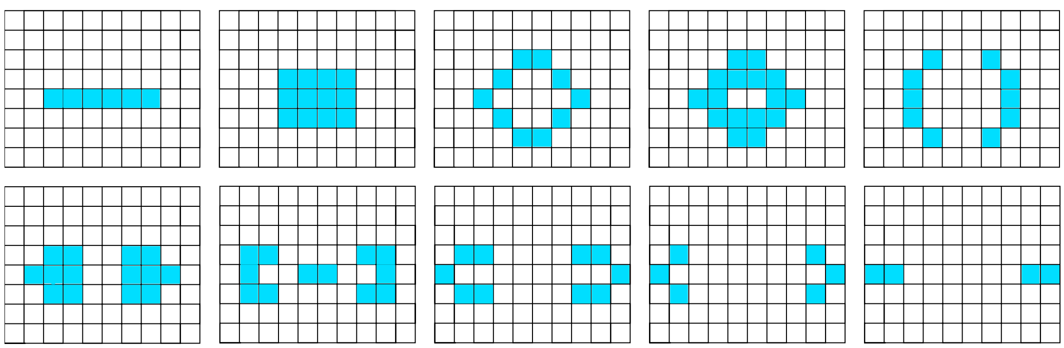
Recitation 4: The Gram-Schmidt Process

Week 4 | *Caltech 2011*

1 Random Question

Consider an infinite $\mathbb{Z} \times \mathbb{Z}$ grid of squares, finitely many of which are marked 1 (alive)and the rest of which are marked 0 (dead). We "iterate" this grid of squares by implementing the following three rules simultaneously:

- If a live cell has < 2 neighbors[1], it dies of loneliness.
- If a live cell has > 3 neighbors, it dies of overcrowding.
- If a dead cell has precisely 3 neighbors, it becomes alive.



Ten consecutive iterations of the rules described above.

Can you find some configuration of cells that, under the above rules, never stops changing? (i.e. in the example above, the system becomes completely empty after the next step. Is this inevitable?)

2 HW comments

- Section average: 80%, or 64/80.
- There were a few, somewhat critical issues that cropped up a lot on this HW, which I'd like to just briefly repeat here:
- In general, *always* state *why* you're doing what you're doing. If you're using a theorem, state it! If you're making a claim about matrices, tell us why it's true! On these HW's, the matrices you get at the end of your work are really the least important parts of the problems; the methods you use to get there are what we really care about.

[1] The **neighbors** of a cell are the 8 cells that surround it in our grid.

- Many matrices do not have inverses! In general, never assume that A^{-1} exists for a general matrix A.
- Also, matrix multiplication does not commute! In other words, there are many matrices A, B such that $AB \neq BA$. Try to avoid accidentally assuming this!
- Similarly, remember how the transpose and inverse operations commute across multiplication: $(AB)^T = B^T A^T$, and $(AB)^{-1} = B^{-1}A^{-1}$.
- Remember your basic logic! If you are asked to prove an if and only if statement, make sure you're proving *two* statements: the "if" part, and the "only if" part! See me if you're confused; I have lectures I gave last quarter that should be helpful in clearing this up.
- In general: please, please, if you're confused, contact me! I'm usually up quite late and at odd hours; if you have HW questions, I far prefer that you ask me than to remain confused and not know what's going on.

3 Gram-Schmidt: The Motivation

When we study vector spaces, we've often found that having a basis is a remarkably useful thing. If we have a basis for a vector space, we really understand it a lot better than when we don't: they allow us to express *all* of the elements in the entire space as just some sum of elements in the basis, which lets us characterize a very big thing (the vector space itself) via some much smaller thing (its basis.)

However, some bases are more useful than others! Take, for example, the standard basis for $\mathbb{R}^n$:

$$\begin{aligned}(1,0,0,\ldots,0) &= \mathbf{e}_1\\ (0,1,0,\ldots,0) &= \mathbf{e}_2\\ &\vdots\\ (0,0,0,\ldots,1) &= \mathbf{e}_n\end{aligned}$$

This basis has two remarkably useful properties:

- All of the elements in this basis have length 1, as $\sqrt{\langle \mathbf{e}_i, \mathbf{e}_i \rangle} = \sqrt{1} = 1$.
- Any two distinct elements in this basis are orthogonal, as $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$.

These two properties are remarkably useful! So useful, in fact, that we give such bases a special name: any basis that has these two properties is called an **orthonormal** basis. (A basis that has just the first property is called a **normal** basis; a basis that has just the second property is called an **orthogonal** basis.)

So: suppose we have some finite-dimensional space V and a basis B for V. Is there a way to turn B into one of these really nice orthonormal bases?

As it turns out: yes!

4 Gram-Schmidt: The Algorithm

To do this, we use the **Gram-Schmidt process**, which we describe here:

- As input, we take in a finite-dimensional space V and a basis B for V. As output, we will create an orthogonal basis U for B, and then make an orthonormal basis E for B using U.
- Suppose that $B = \{\mathbf{b}_1, \ldots \mathbf{b}_n\}$. We create n vectors $\mathbf{u}_1, \mathbf{u}_n$ and n vectors $\{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ from the elements in B inductively, as follows:

$$
\begin{aligned}
\mathbf{u}_1 &= \mathbf{b}_1, & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \\
\mathbf{u}_2 &= \mathbf{b}_2 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{b}_2), & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} \\
\mathbf{u}_3 &= \mathbf{b}_3 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{b}_3) - \mathrm{proj}_{\mathbf{u}_2}(\mathbf{b}_3), & \mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \\
&\vdots & &\vdots \\
\mathbf{u}_n &= \mathbf{b}_n - \sum_{j=1}^{n-1} \mathrm{proj}_{\mathbf{u}_j}(\mathbf{b}_n), & \mathbf{e}_n &= \frac{\mathbf{u}_n}{\|\mathbf{u}_n\|},
\end{aligned}
$$

  where the projection operator $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ is defined as follows:

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u}.$$

- U, the collection of the various $\mathbf{u}_i$'s, clearly has the same span as B, as you can write any of the $\mathbf{b}_i$'s in terms of elements from U (by our above definitions.) As well, you can prove by induction that for any $\mathbf{u}_i, \mathbf{u}_j$ with $i \neq j$, these two vectors are orthogonal: to do this,
  - start by showing that $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$ is 0,
  - use this to show that $\langle \mathbf{u}_1, \mathbf{u}_3 \rangle$ is also 0, and
  - inductively show that $\langle \mathbf{u}_1, \mathbf{u}_k \rangle$, for every k, and then
  - use induction again! (i.e. you're doing a kind of "double-induction) to show that $\langle \mathbf{u}_j, \mathbf{u}_k \rangle$ is 0 for any $j \neq k$. i.e. induct on j and then on k: the three steps above have given you your base cases. See me if you'd like to see a full proof of this!
- Given this, we're done – we've shown that U is an orthogonal basis for V, and thus (because E is just U with all of its vectors normalized to length 1) we know that E is an orthonormal basis for V.

5 Gram-Schmidt: The Example

To illustrate how this goes, we study an example:

Example. Use Gram-Schmidt to turn the basis $\{(1,1,0),(1,0,1),(0,1,1)\}$ for $\mathbb{R}^3$ into an orthonormal basis.

Proof. So: via the Gram-Schmidt algorithm, we define the vectors $\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3$ as follows:

$$\begin{aligned}
\mathbf{u}_1 &= (1,1,0),\\
\mathbf{u}_2 &= (1,0,1) - \mathrm{proj}_{(1,1,0)}\left((1,0,1)\right)\\
&= (1,0,1) - \frac{\langle (1,1,0),(1,0,1)\rangle}{\langle (1,1,0),(1,1,0)\rangle}\cdot(1,1,0)\\
&= (1,0,1) - \frac{1}{2}\cdot(1,1,0)\\
&= \left(\frac{1}{2},-\frac{1}{2},0\right)\\
\mathbf{u}_3 &= (0,1,1) - \mathrm{proj}_{(1,1,0)}\left((0,0,1)\right) - \mathrm{proj}_{\left(\frac{1}{2},-\frac{1}{2},0\right)}\left((0,1,1)\right)\\
&= (0,1,1) - \frac{\langle (1,1,0),(0,1,1)\rangle}{\langle (1,1,0),(1,1,0)\rangle}\cdot(1,1,0) - \frac{\langle \left(\frac{1}{2},-\frac{1}{2},0\right),(0,1,1)\rangle}{\langle \left(\frac{1}{2},-\frac{1}{2},0\right),\left(\frac{1}{2},-\frac{1}{2},0\right)\rangle}\cdot\left(\frac{1}{2},-\frac{1}{2},0\right)\\
&= (0,1,1) - \frac{1}{2}\cdot(1,1,0) - \frac{1/2}{3/2}\cdot\left(\frac{1}{2},-\frac{1}{2},0\right)\\
&= (0,1,1) - \left(\frac{1}{2},\frac{1}{2},0\right) - \left(\frac{1}{6},-\frac{1}{6},\frac{1}{3}\right)\\
&= \left(-\frac{2}{3},\frac{2}{3},\frac{2}{3}\right).
\end{aligned}$$

These three vectors then form an orthogonal basis U for $\mathbb{R}^3$. To turn them into an orthonormal basis, we divide them by their length:

$$\begin{aligned}
\mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}\\
&= \frac{1}{\sqrt{\langle (1,1,0),(1,1,0)\rangle}}(1,1,0)\\
&= \frac{1}{\sqrt{2}}(1,1,0),\\
\mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}\\
&= \frac{1}{\sqrt{\langle \left(\frac{1}{2},-\frac{1}{2},0\right),(\left(\frac{1}{2},-\frac{1}{2},0\right)\rangle}}\left(\frac{1}{2},-\frac{1}{2},0\right)\\
&= \frac{1}{\sqrt{3/2}}\left(\frac{1}{2},-\frac{1}{2},0\right),
\end{aligned}$$

$$
\begin{aligned}
\mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \\
&= \frac{1}{\sqrt{\langle \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right), \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right) \rangle}} \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right) \\
&= \frac{1}{\sqrt{4/3}} \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right).
\end{aligned}
$$

The vectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ then form an orthonormal basis for $\mathbb{R}^3$, which is what we sought.

6 Gram-Schmidt: The Applications

Gram-Schmidt has a number of really useful applications: here are two quick and elegant results.

Proposition 1 *Suppose that V is a finite-dimensional vector space with basis $\{\boldsymbol{b}_1 \ldots \boldsymbol{b}_n\}$, and $\{\boldsymbol{u}_1, \ldots \boldsymbol{u}_n\}$ is the orthogonal (not orthonormal!) basis that the Gram-Schmidt process creates from the $\boldsymbol{b}_i$'s.*

Let U denote the matrix with rows given by the vectors $\{\boldsymbol{u}_1, \ldots \boldsymbol{u}_n\}$, and let B denote the matrix with rows given by the matrix $\{\boldsymbol{b}_1, \ldots \boldsymbol{b}_n\}$. Then, we have that

$$
B = \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
\frac{\langle \boldsymbol{b}_2, \boldsymbol{u}_1 \rangle}{\langle \boldsymbol{u}_1, \boldsymbol{u}_1 \rangle} & 1 & 0 & 0 & \ldots & 0 \\
\frac{\langle \boldsymbol{b}_3, \boldsymbol{u}_1 \rangle}{\langle \boldsymbol{u}_2, \boldsymbol{u}_1 \rangle} & \frac{\langle \boldsymbol{b}_3, \boldsymbol{u}_2 \rangle}{\langle \boldsymbol{u}_2, \boldsymbol{u}_2 \rangle} & 1 & 0 & \ldots & 0 \\
\frac{\langle \boldsymbol{b}_4, \boldsymbol{u}_1 \rangle}{\langle \boldsymbol{u}_1, \boldsymbol{u}_1 \rangle} & \frac{\langle \boldsymbol{b}_4, \boldsymbol{u}_2 \rangle}{\langle \boldsymbol{u}_2, \boldsymbol{u}_2 \rangle} & \frac{\langle \boldsymbol{b}_4, \boldsymbol{u}_3 \rangle}{\langle \boldsymbol{u}_3, \boldsymbol{u}_3 \rangle} & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\langle \boldsymbol{b}_n, \boldsymbol{u}_1 \rangle}{\langle \boldsymbol{u}_1, \boldsymbol{u}_1 \rangle} & \frac{\langle \boldsymbol{b}_n, \boldsymbol{u}_2 \rangle}{\langle \boldsymbol{u}_2, \boldsymbol{u}_2 \rangle} & \frac{\langle \boldsymbol{b}_n, \boldsymbol{u}_3 \rangle}{\langle \boldsymbol{u}_3, \boldsymbol{u}_3 \rangle} & \frac{\langle \boldsymbol{b}_n, \boldsymbol{u}_4 \rangle}{\langle \boldsymbol{u}_4, \boldsymbol{u}_4 \rangle} & \ldots & 1
\end{bmatrix} \cdot U
$$

Proof. Just solve the equations given by the Gram-Schmidt process for the b_i entries: this tells you which combinations of the U rows will yield an element of B.

The above result may not look too interesting now, but when we get to eigenvalues and change-of-basis matrices, it will come in handy.

Another nice result, whose utility is much more obvious, is the following:

Proposition 2 *For V a vector space, U a finite-dimensional subspace of V, $B = \{\boldsymbol{b}_1 \ldots \boldsymbol{b}_n\}$ an orthogonal basis for U, and any vector $\boldsymbol{x}$ in V, we define the **orthogonal projection** of $\boldsymbol{x}$ onto U as the following vector in U:*

$$
\mathrm{proj}_U(\boldsymbol{x}) = \sum_{i=1}^{n} \mathrm{proj}_{\mathbf{b}_i}(\mathbf{x}).
$$

This vector is the closest vector in U to $\boldsymbol{x}$.

We reserve the proof of this statement for class, and instead focus on an example of its use:

Example. Let $V = C[0,1]$, the space of all continuous functions on the interval $[0,1]$; as the sum of any two continuous functions is again a continuous function, and scalar multiples of continuous functions is still continuous, this is clearly a vector space. Let U denote all of the polynomials with real coefficients and degree ≤ 2; this is clearly a subspace of V, with basis $\{1, x, x^2\}$.

What is the closest element in U to the element e^x in V?

Proof. So: first, recall that the inner-product on the space of all real-valued continuous functions from 0 to 1 is defined as the following:

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx.$$

We seek to find the closest degree $\leq$ 2-polynomial to e^x with respect to this idea of distance: in other words, we're trying to find a polynomial that stays close to e^x on *all* of [0,1]. (This contrasts nicely with Taylor series, which only try to approximate a given function at a single point.)

To do this, we need to do just two things:

1. Find an orthogonal basis for U, and
2. Use this to find the orthogonal projection of e^x onto U.

To do the first, we apply the Gram-Schmidt process to the basis $\{1, x, x^2\}$:

$$\begin{aligned}
\mathbf{u}_1 &= 1,\\
\mathbf{u}_2 &= x - \mathrm{proj}_1(x)\\
&= x - \frac{\langle 1, x\rangle}{\langle 1, 1\rangle}\cdot 1\\
&= x - \frac{\int_0^1 1\cdot x dx}{\int_0^1 1\cdot 1 dx}\cdot 1\\
&= x - \frac{x^2/2\Big|_0^1}{x\Big|_0^1}\cdot 1\\
&= x - \frac{1/2}{1}\cdot 1\\
&= x - \frac{1}{2},\\
\mathbf{u}_3 &= x^2 - \mathrm{proj}_1(x^2) - \mathrm{proj}_{x-\frac{1}{2}}(x^2)\\
&= x^2 - \frac{\langle 1, x^2\rangle}{\langle 1, 1\rangle}\cdot 1 - \frac{\langle x-\frac{1}{2}, x^2\rangle}{\langle x-\frac{1}{2}, x-\frac{1}{2}\rangle}\cdot (x - \frac{1}{2})
\end{aligned}$$

$$
\begin{aligned}
&= x^2 - \frac{\int_0^1 1 \cdot x^2 dx}{\int_0^1 1 \cdot 1 dx} \cdot 1 - \frac{\int_0^1 (x-\frac{1}{2}) \cdot x^2 dx}{\int_0^1 (x-\frac{1}{2})^2 \cdot 1 dx} \cdot (x - \frac{1}{2}) \\
&= x^2 - \frac{x^3/3\Big|_0^1}{x\Big|_0^1} \cdot 1 - \frac{x^4/4 - x^3/6\Big|_0^1}{x^3/3 - x^2/2 + x/4\Big|_0^1} \cdot (x - \frac{1}{2}) \\
&= x^2 - \frac{1/3}{1} \cdot 1 - \frac{1/4 - 1/6}{1/3 - 1/2 + 1/4} \cdot (x - \frac{1}{2}) \\
&= x^2 - x + \frac{1}{6}.
\end{aligned}
$$

Therefore, an orthogonal basis for U is $\{1, x-\frac{1}{2}, x^2-x+\frac{1}{6}\}$. Using this, we can calculate the projection of e^x onto U:

$$
\text{proj}_U(e^x) = \text{proj}_1(e^x) + \text{proj}_{x-\frac{1}{2}}(e^x) = \text{proj}_{x^2-x+\frac{1}{6}}(e^x),
$$

where

$$
\begin{aligned}
\text{proj}_1(e^x) &= \frac{\langle 1, e^x \rangle}{\langle 1, 1 \rangle} \cdot 1 \\
&= \frac{\int_0^1 e^x}{1} \cdot 1 \\
&= e - 1, \\
\text{proj}_{x-\frac{1}{2}}(e^x) &= \frac{\langle x - \frac{1}{2}, e^x \rangle}{\langle x - \frac{1}{2}, x - \frac{1}{2} \rangle} \cdot (x - \frac{1}{2}) \\
&= \frac{\int_0^1 xe^x - e^x/2 dx}{\int_0^1 (x-\frac{1}{2})^2 \cdot 1 dx} \cdot (x - \frac{1}{2}) \\
&= \frac{xe^x - e^x - e^x/2\Big|_0^1}{x^3/3 - x^2/2 + x/4\Big|_0^1} \cdot (x - \frac{1}{2}) \\
&= \frac{3/2 - e/2}{1/12}(x - \frac{12}{)} \\
&= (3-e)(6x-3), \text{ and} \\
\text{proj}_{x^2-x+\frac{1}{6}}(e^x) &= \frac{\langle x^2 - x + \frac{1}{6}, e^x \rangle}{\langle x^2 - x + \frac{1}{6}, x^2 - x + \frac{1}{6} \rangle} \cdot (x^2 - x + \frac{1}{6}) \\
&= \frac{\int_0^1 x^2e^x - xe^x + e^x/6 dx}{\int_0^1 x^4 - 2x^3 - \frac{2}{3}x^2 - \frac{1}{3}x + \frac{1}{36} dx} \cdot (x^2 - x + \frac{1}{6})
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{x^2e^x - 2xe^x + 2e^x - xe^x + e^x + e^x/6\Big|_0^1}{\frac{1}{5}x^5 - \frac{1}{2}x^4 - \frac{2}{9}x^3 - \frac{1}{6}x^2 + \frac{1}{36}x\Big|_0^1} \cdot (x^2 - x + \frac{1}{6}) \\
&= \frac{\frac{7}{6}e - \frac{13}{6}}{\frac{1}{5} - \frac{1}{2} - \frac{2}{9} - \frac{1}{6} + \frac{1}{36}} \cdot (x^2 - x + \frac{1}{6}) \\
&= \frac{390 - 210e}{119} \cdot (x^2 - x + \frac{1}{6}).
\end{aligned}
$$

and therefore, we have that the closest degree-2 polynomial to e^x on the interval $[0, 1]$ is

$$
\frac{(390 - 210e)}{119}x^2 + \left((18 - 3e) - \frac{(390 - 210e)}{119}\right)x + \left(4e - 10 - \frac{(390 - 210e)}{714}\right) \cdot 1.
$$

Cool, right?