Spaces of Functions

Earlier in the course, we said a **vector space** was a collection S of objects along with rules for

- how to add any two objects in S, and
- how to rescale objects in S by any scalar.

We demanded that S be closed under these operations of addition and scalar multiplication, and that eight algebraic properties (among them, the existence of an additive identity, or *zero*, in S) hold.

While the Euclidean spaces $\mathbb{R}^n$ must surely have served as inspiration for this abstracted definition, we found some other sets with familiar addition and scalar multiplication operations that met all the criteria. Function spaces, such as these mentioned below, are among them.

- The real-valued functions defined on domain $[0, 1]$.
- The real-valued functions defined on domain $(-\infty, \infty)$.
- The continuous, real-valued functions defined on domain (a, b), often referred to using the symbol $\mathscr{C}(a, b)$.
- The k-times continuously differentiable, real-valued functions defined on domain (a, b), often referred to using the symbol $\mathscr{C}^k(a, b)$.

Let's do a bit of comparison between a space like $\mathscr{C}(a, b)$ and $\mathbb{R}^n$.

Components of the vectors. A vector in $\mathbb{R}^3$ has 3 components. It is natural to ask the value of the 2^{nd} component of $\mathbf{u} \in \mathbb{R}^n$. The value of the 2^{nd} component of $(5, -1, 4)$ is (-1).

For a function defined on (a, b), there are infinitely-many components, one for each $x \in (a, b)$. The value of the component corresponding to some x is $f(x)$.

Additive identity. In $\mathbb{R}^n$ the additive identity is the element for which each of its n components has value 0. This vector, added to any other vector in $\mathbb{R}^n$, leaves the latter intact. In function spaces, the element you can add to any other yet leave it unchanged is the constant function $f(x) = 0$ (i.e., for each input $x \in (a, b)$, the output is 0).

Linear independence. In $\mathbb{R}^n$, we say a collection $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is linearly independent if the only linear combination

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n$$

that produces the zero vector is the one for which every one of the weights $c_1, \ldots, c_n$ is zero. This notion translates quite easily to function space. If you have a collection $S = \{f_1, f_2, \ldots, f_n\}$ of real-valued functions defined on (a, b) and you wish to produce the zero function on (a, b) as a linear combination

$$c_1 f_1 + c_2 f_2 + \cdots + c_n f_n,$$

the question of linear independence of S again comes down to whether there is any other way to do it besides taking all of the weights $c_1, \ldots, c_n$ to be 0. There are still instances in which collections are linearly dependent, such as $\{1, \cos^2 x, \sin^2 x\}$, since

$$\underline{\text{(-1)}}\, 1 + \underline{\ \ 1\ \ }\, \cos^2 x + \underline{\ \ 1\ \ }\, \sin^2 x \ = \ 0.$$

Notice that the commonality of domain (a, b) for the functions considered is important here. One cannot hope to take a linear combination of $f(x) = x$ and $g(x) = 1/x$ (remembering that the weights must be *scalars*), and produce the zero function defined on $(-\infty, \infty)$, for the equation

$$c_1 x + c_2 \left(\frac{1}{x}\right) \ = \ 0$$

breaks down at $x = 0$. The perceived problem arises only if we are thinking that a function defined on $(-\infty, \infty)$ added with a function defined, say, on $(0, \infty)$, should be able to produce a function defined on $(-\infty, \infty)$. If we think of f, g and the zero function as all defined on the limited domain $(0, \infty)$, there is no problem. One encounters much the same issue if trying to add (or take a linear combination) of vectors from different Euclidean spaces, say $(1, 1, 3)$ and $(2, 1, 5, -1, 1)$, hoping to produce the zero vector.

Dimension. The dimension of a space is the number of elements in a basis of that space. A basis of $\mathbb{R}^n$ is the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, where $\mathbf{e}_1 = (1, 0, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$ $\mathbf{e}_n = (0, 0, 0, \ldots, 1)$. One can see that there are n of these vectors $\mathbf{e}_j$, so that the dimension of $\mathbb{R}^n = n$, the same as the number of components each vector $\mathbf{v} \in \mathbb{R}^n$ has.

Since functions have components, too, one for each $x \in (a, b)$, we might guess (correctly!) that $\mathscr{C}(a, b)$ is of *infinite* dimension. Yet such a space has finite-dimensional subspaces, such as $W = \text{span}(\{1, x, x^2\})$. One can check that as functions defined, say, on $(0, 1)$, 1, x and x^2 are linearly independent. Thus, they form a basis for the 3-dimensional space W consisting of polynomials defined on $(0, 1)$ of degree 2 or less.

Inner products. We calculate the dot (or *inner*) product of two vectors $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ by summing the products of corresponding coordinates:

$$\mathbf{u} \cdot \mathbf{v} \ = \ \sum_{j=1}^{n} u_j v_j.$$

For functions f, g in $\mathscr{C}(a, b)$, the coordinate of f at x is $f(x)$, and the product with the corresponding component of g is $f(x)g(x)$. If we want to sum up all such products, we get

$$\langle f, g \rangle \ := \ \int_a^b f(x)g(x)\, dx. \tag{1}$$

The notation $\langle f, g \rangle$ is preferred over symbols like $f \cdot g$ and $f^{\mathrm{T}} g$ which have been used for vectors in Euclidean space. The inner product $\langle f, g \rangle$ defined in (1) is called the $L^2(a,b)$ **inner product**, or just the L^2 **inner product**.

One might verify (or take for granted) that this definition for an inner product between functions has the various properties we have come to expect from dot products of vectors, namely

	$\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $c \in \mathbb{R}$	$\mathbf{f}, \mathbf{g}, \mathbf{h} \in \mathscr{C}(a,b)$, $c \in \mathbb{R}$
(i)	$\mathbf{v} \cdot \mathbf{v} \geqslant 0$, for all $\mathbf{v}$	$\langle f, f \rangle = \int_a^b \lvert f(x) \rvert^2 \, dx \geqslant 0$ for all f
(ii)	$\mathbf{v} \cdot \mathbf{v} = 0$ implies $\mathbf{v} = \mathbf{0}$	$\langle f, f \rangle = \int_a^b \lvert f(x) \rvert^2 \, dx = 0$ implies $f = 0$
(iii)	$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$	$\langle f, g \rangle = \int_a^b f(x) g(x) \, dx = \langle g, f \rangle$
(iv)	$(c u) \cdot \mathbf{v} = c(\mathbf{v} \cdot \mathbf{u})$	$\langle cf, g \rangle = \int_a^b c f(x) g(x) \, dx = c(\langle f, g \rangle)$
(v)	$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$	$\langle f, g + h \rangle = \int_a^b f(x)[g(x) + h(x)] \, dx = \langle f, g \rangle + \langle f, h \rangle$

Length. One can obtain the length of a vector $\mathbf{v} \in \mathbb{R}^n$ as the square root of its dot product with itself:

$$\|\mathbf{v}\| \; = \; \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

We can extend this idea to give us a function's length, as a member of $\mathscr{C}(a,b)$. That is, we take

$$\|f\| \; = \; \sqrt{\langle f, f \rangle} \; = \; \left[\int_a^b \lvert f(x) \rvert^2 \, dx \right]^{1/2}.$$

Thus, as a function in $\mathscr{C}(0,1)$, the function $f(x) = x^2$ has "length"

$$\|f\| \; = \; \left(\int_0^1 x^4 \, dx \right)^{1/2} \; = \; \frac{1}{\sqrt{5}}.$$

Once again, the domain makes a difference for, as a function in $\mathscr{C}(-5,5)$, the function $f(x) = x^2$ has length

$$\|f\| \; = \; \left(\int_{-5}^5 x^4 \, dx \right)^{1/2} \; = \; 25\sqrt{2}.$$

Orthogonality. We say vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$. Similarly, we can define functions f, $g \in \mathscr{C}(a,b)$ to be orthogonal (over that domain, perhaps not if over a different one) precisely when $\langle f, g \rangle = 0$. For instance, the functions $\cos x$ and $\sin x$ are orthogonal over $(0, \pi)$, since

$$\langle \cos x, \sin x \rangle \; = \; \int_0^\pi \cos x \sin x \, dx \; = \; \frac{1}{2} \int_0^\pi \sin(2x) \, dx \; = \; -\frac{1}{4} \cos(2x) \Big|_0^\pi \; = \; 0.$$

Though they are orthogonal as functions in $\mathscr{C}(0, \pi)$, one should not assume $\sin x$ and $\cos x$ are orthogonal as functions defined on an arbitrary interval (a,b). The power functions 1 and x^2 in $\mathscr{C}(0,1)$ are *not* orthogonal:

$$\langle 1, x^2 \rangle \; = \; \int_0^1 (1)(x^2) \, dx \; = \; \frac{1}{3} x^3 \Big|_0^1 \; = \; \frac{1}{3}.$$

The Gram-Schmidt algorithm in function space

The algorithm is carried out in much the same way in function space as in Euclidean space, modified only to reflect the use of L^2 inner product between functions. So, if we have a collection of linearly independent functions $\{f_1, \ldots, f_n\}$, we produce orthonormal functions $\{q_1, \ldots, q_n\}$ via

$$\begin{aligned} u_1 &= f_1 & \text{then set} \quad q_1 &= \frac{u_1}{\|u_1\|}, \\ u_2 &= f_2 - \frac{\langle f_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 & \text{then set} \quad q_2 &= \frac{u_2}{\|u_2\|}, \\ u_3 &= f_3 - \frac{\langle f_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle f_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2 & \text{then set} \quad q_3 &= \frac{u_3}{\|u_3\|}, \\ &\text{etc.} \end{aligned}$$

Example 1:

We carry out Gram-Schmidt on the independent set $\{1, x, x^2\}$ in $\mathscr{C}(0,1)$.

We set $u_1 = 1$. Then

$$u_2 = x - \frac{\langle x, 1 \rangle}{\langle 1, 1 \rangle} \cdot 1 = x - \frac{1/2}{1} \cdot 1 = x - \frac{1}{2}.$$

As before, orthogonality between u_1 and u_2 is maintained if we rescale either function, and it seems convenient to rescale to $2u_2 = 2x - 1$. So, let us now say

$$u_2 = 2x - 1.$$

Then

$$u_3 = x^2 - \frac{\langle x^2, 1 \rangle}{\langle 1, 1 \rangle} \cdot 1 - \frac{\langle x^2, 2x-1 \rangle}{\langle 2x-1, 2x-1 \rangle} \cdot (2x-1) = x^2 - \frac{1/3}{1} \cdot 1 - \frac{1/6}{1/3} \cdot (2x-1) = x^2 - x + \frac{1}{6}.$$

Again, multiplying u_3 by 6 does not change its orthogonality with u_1 and u_2, so it is all right to take

$$u_3 = 6x^2 - 6x + 1.$$

Now we divide u_1, u_2 and u_3 by their lengths to obtain q_1, q_2 and q_3:

$$\begin{aligned} \|u_1\|^2 &= \int_0^1 1^2 \, dx = 1 & \Rightarrow \quad q_1 &= 1. \\ \|u_2\|^2 &= \int_0^1 (2x-1)^2 \, dx = \frac{1}{3} & \Rightarrow \quad q_2 &= \sqrt{3}(2x-1). \\ \|u_3\|^2 &= \int_0^1 (6x^2 - 6x + 1)^2 \, dx = \frac{1}{5} & \Rightarrow \quad q_2 &= \sqrt{5}(6x^2 - 6x + 1). \end{aligned}$$

■

The closest function in a function subspace

In what follows, we will do our work in the context of this example: finding the closest (as measured by $L^2(0,1)$ distance) quadratic function to $f(x) = e^x$. This is *not* going to come down to doing the same thing as was done in Example 3 on p. 226 in our text, as we shall see.

Proceeding naively, one would look to choose coefficients c, b, and a so that

$$p(x) \;=\; c + bx + ax^2 \;=\; \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} \;=\; f(x).$$

This appears to be a 1-by-3 matrix multiplied with a 3-by-1, but the situation is far more complex. We really want equality at each $x \in (0,1)$. That is, we are looking for equality

$$\begin{aligned}
\text{at } x = 1/2: &\quad c + b(0.5) + a(0.5)^2 \;=\; f(0.5)\\
\text{at } x = 1/3: &\quad c + b(0.\overline{3}) + a(0.\overline{3})^2 \;=\; f(0.\overline{3})\\
\text{at } x = 2/3: &\quad c + b(0.\overline{6}) + a(0.\overline{6})^2 \;=\; f(0.\overline{6})\\
\text{at } x = 1/4: &\quad c + b(0.25) + a(0.25)^2 \;=\; f(0.25)\\
\text{at } x = 3/4: &\quad c + b(0.75) + a(0.75)^2 \;=\; f(0.75)
\end{aligned}$$

and so on. There are infinitely many x-values in $(0,1)$, and requiring equality at each one produces a matrix with infinitely many rows; I've listed several below:

$$\begin{bmatrix} 1 & 0.5 & (0.5)^2 \\ 1 & 0.\overline{3} & (0.\overline{3})^2 \\ 1 & 0.\overline{6} & (0.\overline{6})^2 \\ 1 & 0.25 & (0.25)^2 \\ 1 & 0.75 & (0.75)^2 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \begin{bmatrix} e^{0.5} \\ e^{0.\overline{3}} \\ e^{0.\overline{6}} \\ e^{0.25} \\ e^{0.75} \\ \vdots \end{bmatrix}. \tag{2}$$

This is like a *seriously* overdetermined vector problem $\mathbf{Ax} = \mathbf{b}$; $\mathbf{A}$ has an (uncountably) infinite number of rows, and 3 columns. (Can $\mathbf{A}$ still be called a *matrix*?) Most functions f, including the exponential function e^x, when used to generate $\mathbf{b}$, result in a vector that is not in the "column space" of such an $\mathbf{A}$. Note the similarities and the differences between this problem and the one posed in Worked Example **4.3 B** on p. 228. In both problems, it is highly unlikely that a right-hand side vector $\mathbf{b}$ is in the column space of $\mathbf{A}$, making both problems unsolvable. But in **4.3 B** we seek equality at a limited (finite) number of points, whereas here we seek equality at *every* $x \in (0,1)$.

In the theme of Chapter 4, we seek to project e^x into $W = \text{span}(\{1, x, x^2\})$ by first solving normal equations $\mathbf{A}^{\mathrm{T}}\mathbf{Ax} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$. It is possible to think one's way to the normal equations using the $\mathbf{A}$, $\mathbf{b}$ of Equation (2) above. I will take a different tack, however, one that Strang used in Section 4.3 when finding a line to best fit data points, that employs calculus. What we seek to do is minimize the squared $L^2(0,1)$ distance between $f(x) = e^x$ and a quadratic polynomial $p(x) = c + bx + ax^2$ by choosing the best coefficients. That is, we want to make

$$\|f(x) - c - bx - ax^2\|^2 \;=\; \int_0^1 [f(x) - c - bx - ax^2]^2\,dx$$

as small as possible. We define

$$G(a,b,c) \;:=\; \|f(x) - c - bx - ax^2\|^2 \;=\; \int_0^1 [f(x) - c - bx - ax^2]^2\,dx,$$

a function of three variables a, b, c (not x). A necessary condition for G to reach a minimum is that its partial derivatives be zero:

$$\begin{aligned}
0 &= \frac{\partial G}{\partial c} = -2\int_0^1 [f(x) - c - bx - ax^2]\cdot 1\,dx = -2\langle f(x) - p(x), 1\rangle,\\
0 &= \frac{\partial G}{\partial b} = -2\int_0^1 [f(x) - c - bx - ax^2]x\,dx = -2\langle f(x) - p(x), x\rangle,\\
0 &= \frac{\partial G}{\partial a} = -2\int_0^1 [f(x) - c - bx - ax^2]x^2\,dx = -2\langle f(x) - p(x), x^2\rangle.
\end{aligned}$$

Using properties of inner products, we can manipulate these three equations to say

$$\begin{aligned}
\langle f(x), 1\rangle &= \langle p(x), 1\rangle = \langle c + bx + ax^2, 1\rangle = \langle 1,1\rangle c + \langle x,1\rangle b + \langle x^2,1\rangle a\\
\langle f(x), x\rangle &= \langle p(x), x\rangle = \langle c + bx + ax^2, x\rangle = \langle 1,x\rangle c + \langle x,x\rangle b + \langle x^2,x\rangle a\\
\langle f(x), x^2\rangle &= \langle p(x), x^2\rangle = \langle c + bx + ax^2, x^2\rangle = \langle 1,x^2\rangle c + \langle x,x^2\rangle b + \langle x^2,x^2\rangle a.
\end{aligned}$$

Written as a vector equation, this system of 3 equations in the unknowns a, b, and c takes the form

$$\begin{bmatrix} \langle 1,1\rangle & \langle x,1\rangle & \langle x^2,1\rangle \\ \langle 1,x\rangle & \langle x,x\rangle & \langle x^2,x\rangle \\ \langle 1,x^2\rangle & \langle x,x^2\rangle & \langle x^2,x^2\rangle \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \begin{bmatrix} \langle e^x,1\rangle \\ \langle e^x,x\rangle \\ \langle e^x,x^2\rangle \end{bmatrix}. \tag{3}$$

If Equation (2) is to be thought of as $\mathbf{Ax} = \mathbf{b}$, then Equation (3) is its corresponding normal equations, $\mathbf{A}^{\mathrm{T}}\mathbf{Ax} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$. Computing all these inner products, the resulting vector equation is

$$\begin{bmatrix} 1.0 & 0.5 & 0.\overline{3} \\ 0.5 & 0.\overline{3} & 0.25 \\ 0.\overline{3} & 0.25 & 0.2 \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \begin{bmatrix} 1.71828 \\ 1.0 \\ 0.71828 \end{bmatrix}.$$

Octave has a function, `hilb()`, that produces (square) matrices where the entry in position (i, j) is $1/(i + j - 1)$, just as is the case here. So, we obtain the coefficients:

```
octave:947> inv(hilb(3))*[1.71828;  1;  0.71828]
ans =

   1.01292
   0.85152
   0.83880
```

Our best/closest polynomial (with some built-in roundoff error), in the sense of minimizing $L^2(0,1)$ distance, to e^x is

$$p(x) \;=\; \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 1.01292 \\ 0.85152 \\ 0.8388 \end{bmatrix} \;=\; 1.01292 + 0.85152x + 0.8388x^2.$$

Let us not forget that the space W of quadratic polynomials defined on $(0,1)$ has an orthogonal basis we took pains to find in the last section. Were we to use that basis, $\{q_1, q_2, q_3\}$, with

$$q_1(x) = 1, \qquad q_2(x) = \sqrt{3}(2x-1), \qquad q_3(x) = \sqrt{5}(6x^2 - 6x + 1),$$

instead of $\{1, x, x^2\}$, it seems our work might be front-loaded with performing the Gram-Schmidt process, only to be easier now. Indeed, using this basis, the normal equations become

$$\begin{bmatrix} \langle q_1, q_1\rangle & \langle q_2, q_1\rangle & \langle q_3, q_1\rangle \\ \langle q_1, q_2\rangle & \langle q_2, q_2\rangle & \langle q_3, q_2\rangle \\ \langle q_1, q_3\rangle & \langle q_2, q_3\rangle & \langle q_3, q_3\rangle \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} \;=\; \begin{bmatrix} \langle e^x, q_1\rangle \\ \langle e^x, q_2\rangle \\ \langle e^x, q_3\rangle \end{bmatrix}.$$

The orthonormality of the basis $\{q_1, q_2, q_3\}$ immediately gives us that the coefficient matrix on the left is $\mathbf{I}_3$. Computing the inner products on the right-hand side yields the answer

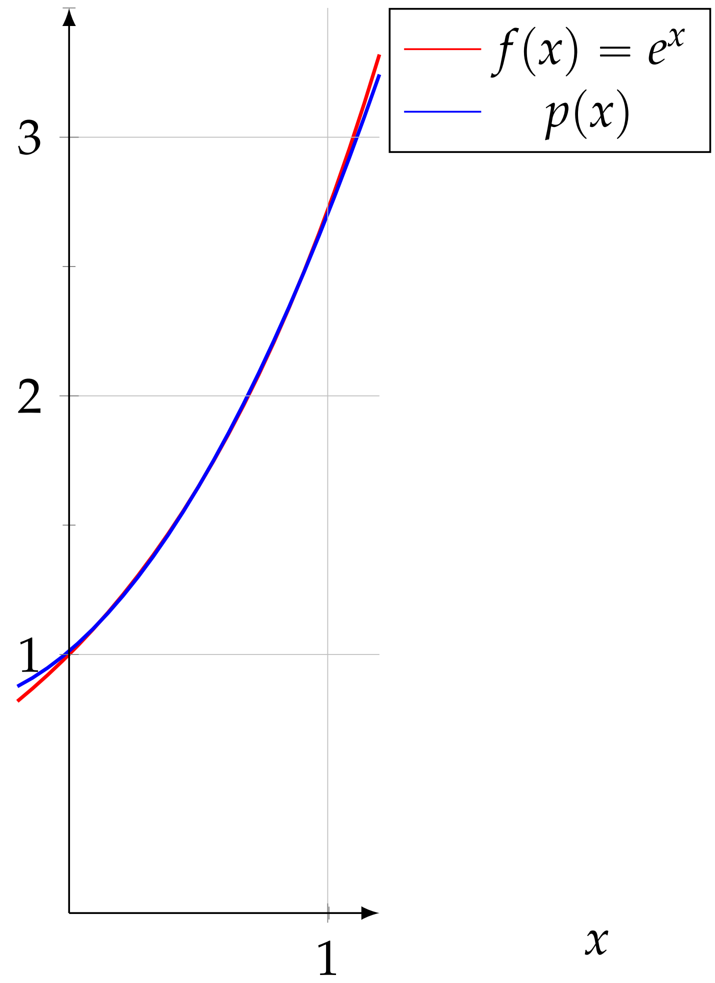
$$\begin{bmatrix} c \\ b \\ a \end{bmatrix} \;=\; \begin{bmatrix} 1.71828 \\ 0.48795 \\ 0.06255 \end{bmatrix}.$$

The resulting polynomial is

$$\begin{aligned} p(x) \;&=\; \begin{bmatrix} 1 & \sqrt{3}(2x-1) & \sqrt{5}(6x^2 - 6x + 1) \end{bmatrix} \begin{bmatrix} 1.71828 \\ 0.48795 \\ 0.06255 \end{bmatrix} \\ &=\; 1.71828 + \sqrt{3}(0.48795)(2x-1) + \sqrt{5}(0.06255)(6x^2 - 6x + 1) \;\;= \cdots \\ &=\; 1.013 + 0.85111x + 0.8392x^2, \end{aligned}$$

largely in agreement with our findings above.

We finish off this example by plotting the two functions, $f(x) = e^x$ and $p(x)$, below.


$f(x) = e^x$
$p(x)$
3
2
1
1
x