A first look at basis expansions

In the last lecture, we looked at the sampling theorem from the point of view of frequency-domain transformations. This is a very good way to understand it in the context of classical signal processing. We can, however, think of it in another way: as an orthobasis expansion in a Hilbert space. This is a powerful viewpoint, as it will allow us to generalize what it means to **discretize** a continuous time signal. It will also give us a unified framework of treating all different types of signals (continuous, discrete, infinite length, time-limited, etc.)

We will start with some concrete examples, and then develop the general framework. Essentially, we will be concerned with extending and abstracting the key concepts from **linear algebra**:

- linear subspaces
- norms
- bases / change of bases
- inner products / orthogonality / projections
- linear operators (matrices in finite dimensions)
- eigenvalues / singular values

Good sources for the material in the rest of Section I include:

- Moon and Stirling, Chapter 2
- G. Strang, "Linear Algebra and its Applications"
- N. Young, "An Introduction to Hilbert Space"
- Naylor and Sell, "Linear Operator Theory in Engineering and Science"

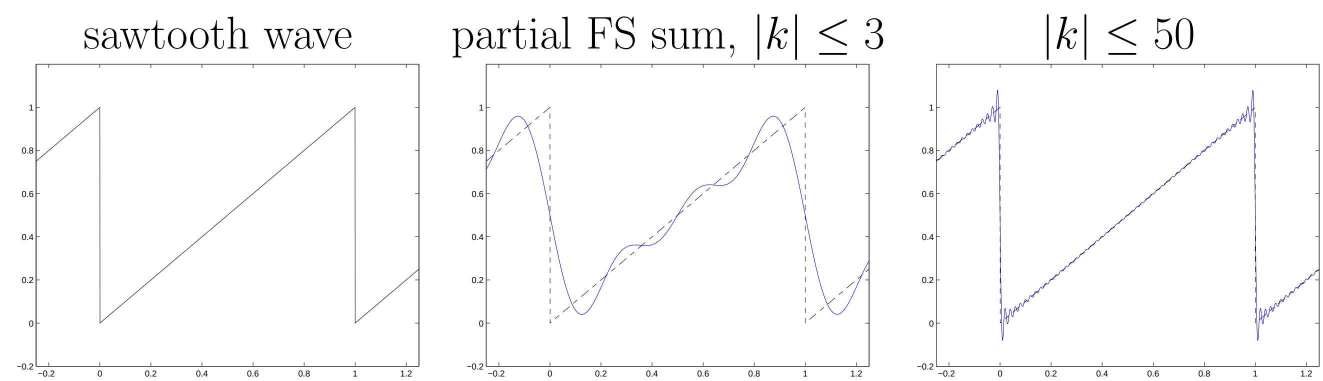
Example: Fourier series

Recall that any periodic signal can be written as a (possibly infinite) superposition of **harmonic sinusoids**. If $x(t)$ has period T, we can write

$$x(t) = \sum_{k=-\infty}^{\infty} \alpha_k \, e^{j2\pi kt/T}, \tag{1}$$

$$\text{where} \quad \alpha_k = \frac{1}{T}\int_{-T/2}^{T/2} x(t)\, e^{-j2\pi kt/T} \, \mathrm{d}t. \tag{2}$$

(The integral above can be computed over any interval of length T.) The two equations above are another example of a reproducing formula — (2) shows how to systematically take a signal and map it into a discrete list of numbers, while (1) shows how to take that list of numbers and synthesize the signal.



Equivalently, we can think of the Fourier series as building up a function that is time-limited to $[-T/2, T/2]$. That every ("reasonable") function can be represented this way is a deep result in mathematics, which we will talk a little more about later. But it is **mathematically equivalent to the sampling theorem**, we just switch the roles of time and frequency.

To see this, suppose that $x(t)$ is zero outside of $[-T/2, T/2]$, so (1) is building it up only inside this interval. Then its Fourier transform is

$$X(j\Omega) = \int_{-T/2}^{T/2} x(t)\, e^{-j\Omega t}\ \mathrm{d}t.$$

Notice that the Fourier series coefficients α_k in (2) are samples of the Fourier transform spaced $2\pi/T$ apart and scaled by $1/T$:

$$\alpha_k = \frac{1}{T}\, X\left(j\frac{2\pi k}{T}\right).$$

Now we can write the Fourier transform as a combination of samples α_k:

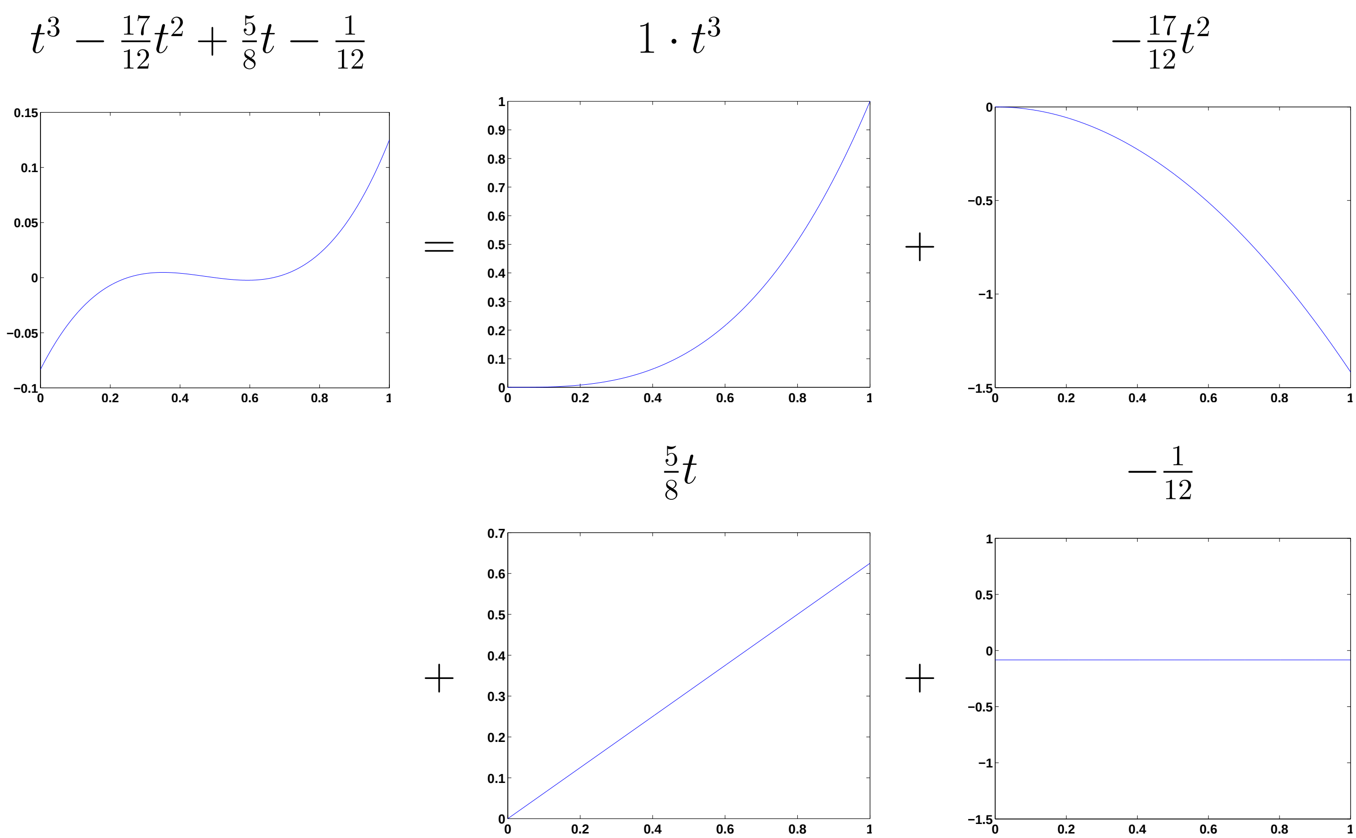
$$\begin{aligned}
X(j\Omega) &= \int_{-T/2}^{T/2} \sum_{k=-\infty}^{\infty} \alpha_k\, e^{j2\pi kt/T}\, e^{-j\Omega t}\ \mathrm{d}t \\
&= \sum_{k=-\infty}^{\infty} \alpha_k \int_{-T/2}^{T/2} e^{j(2\pi k/T-\Omega)t}\ \mathrm{d}t \\
&= \sum_{k=-\infty}^{\infty} \alpha_k\, \frac{2T\sin(\Omega T/2-\pi k)}{\Omega T-2\pi k} \\
&= \sum_{k=-\infty} X\left(j\frac{2\pi k}{T}\right) g_{2\pi/T}(\Omega-2\pi k/T),
\end{aligned}$$

where as in the last set of notes $g_{2\pi/T}(\Omega)$ is a sinc function. This is exactly the same reproducing formula we had for the Shannon-Nyquist sampling theorem. Here it says that the Fourier transform of a signal which is time-limited to T can be reconstructed from samples taken $2\pi/T$ in frequency.

Example: Taylor series

It is almost too obvious that any mth order polynomial can be written as a super position of the $m+1$ functions $1, t, t^2, \ldots, t^m$. (Indeed, this is pretty much the definition of polynomial.)

For example:



More generally, **analytic** functions can be re-written as "infinite degree" polynomials using a Taylor series. For instance, on the interval

$[-1/2, 1/2]$, we can write

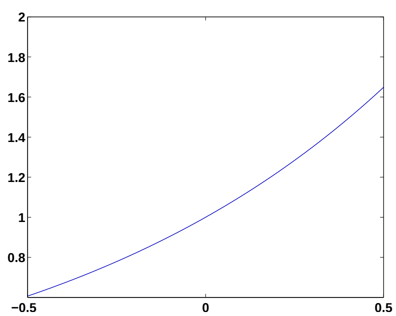
$$e^t = \sum_{k=0}^{\infty} \alpha_k \cdot t^k, \quad \text{where } \alpha_k = \frac{1}{k!},$$

$$\log(1+t) = \sum_{k=0}^{\infty} \alpha_k \cdot t^k \quad \text{where } \alpha_k = \frac{(-1)^{k+1}}{k},$$

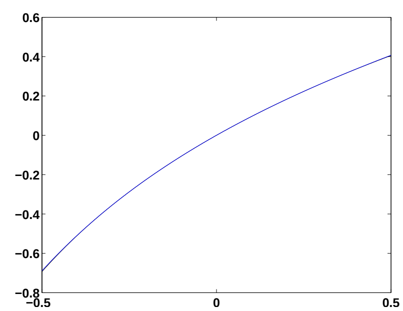
$$\sin(2\pi t) = \sum_{k=0}^{\infty} \alpha_k \cdot t^k \quad \text{where } \alpha_k = \begin{cases} \frac{(-1)^{(k+3)/2}(2\pi)^{k+1}}{(k+1)!} & k \text{ odd} \\ 0 & k \text{ even} \end{cases}$$

Here are the three examples above with the series truncated to the first six terms:

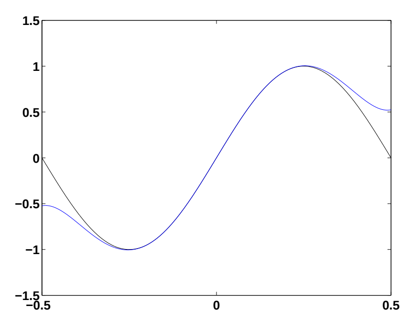
$e^t \approx \sum_{k=1}^{6} \frac{1}{k!} t^k$



$\log(1+t) \approx \sum_{k=1}^{6} \frac{(-1)^{k+1}}{k} t^k$



$\sin(2\pi t) \approx t - \frac{2\pi}{6} t^3 + \frac{8\pi^3}{120} t^5$



The exp and log examples are pretty much spot-on with only six terms, while the sin example is still suffering a little on the edges.

Well-defined "infinite degree" polynomials are called **analytic functions**[1]. For these functions, there is a systematic way of computing the expansion coefficients to represent $\boldsymbol{x}$ on some interval (say $[-1/2, 1/2]$ again),

$$x(t) = \sum_{k=0}^{\infty} \alpha_k \cdot t^k, \quad \text{where} \quad \alpha_k = \frac{x^{(k)}(0)}{k!},$$

where $\boldsymbol{x}^{(k)}$ is the k-th derivative of $\boldsymbol{x}$.

It is important to realize that Taylor series is not the only way to build up a function as a sum of polynomials, and despite its convenience, it has a few unsatisfying properties (e.g. there are infinitely differentiable functions whose Taylor series converges, but does not equal the original function anywhere). Moreover, it is unclear how to use Taylor series for signals that only have a small number of derivatives.

Example: Lagrange polynomials

The sampling theorem from the last set of notes showed that we can build-up a bandlimited signal from a (possibly infinite) superposition of sinc functions; for example, if $x(t)$ is bandlimited to $B = \pi$, then

$$x(t) = \sum_{k=-\infty}^{\infty} \alpha_k g(t-k), \quad g(t) = \frac{\sin(\pi t)}{\pi t},$$

and the expansion coefficient are simply samples of $x(t)$: $\alpha_k = x(k)$. We are able to reproduce $x(t)$ in this example because it adheres to known model: its Fourier transform is zero outside of $[-\pi, \pi]$.

[1]More technically, a function $x(t)$ is called analytic on an interval $[a, b]$ if for any $t_0 \in [a, b]$, the infinite sum $\sum_{k \geq 0} a_k (t - t_0)^k$ converges to $x(t)$ for some choice of $\{a_k\}$.

It might be that we have a different model for the continuous-time signal $x(t)$. One alternative model might be that $x(t)$ is a polynomial. In fact, given a finite number $M+1$ of samples of $x(t)$, there is always an M-th order polynomial that passes through all of them — another way of saying this is that an M-th order polynomial can be reproduced from $M+1$ samples. If $x(t)$ is an M-th order polynomial, then using samples $x(0), x(T), \ldots, x(MT)$, we can write

$$x(t) = \sum_{k=0}^{M} \alpha_k \, p_k(t)$$

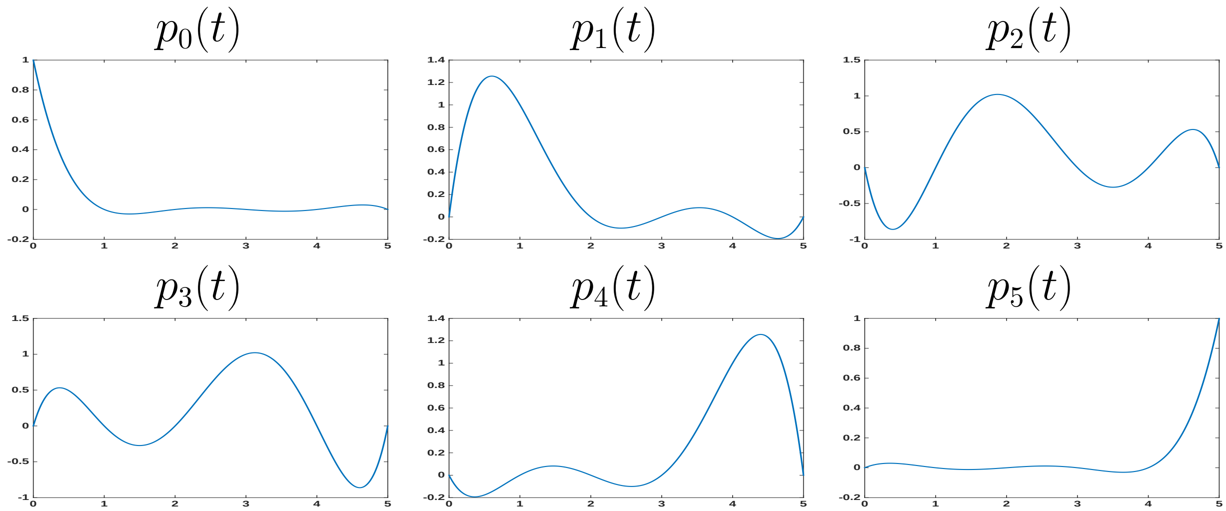
where $\alpha_k = x(kT)$, and

$$p_k(t) = \prod_{\substack{0 \leq m \leq M \\ m \neq k}} \frac{t - mT}{(k-m)T}.$$

It should be clear from the expression above that each $p_k(t)$ is a different M-th order polynomial, and

$$p_k(nT) = \begin{cases} 1, & n = k \\ 0, & n \neq k. \end{cases}$$

Given the samples $x(kT)$, moving to continuous-time signal $x(t)$ is called **Lagrange interpolation**. Here are the 6 basis functions $p_k(t)$ for $M = 5$:



One problem with Lagrange polynomials is that they are extremely unstable outside of the interval $[0, M]$ — they diverge very quickly to either ∞ or $-\infty$ (as all polynomials must do).

Example: Splines

A more stable way to interpolate between a sequence of discrete points is by using a **polynomial spline**. Given a sequence of locations $t_1, t_2, \ldots, t_K$ and function values at those locations $v_{t_1}, v_{t_2}, \ldots, v_{t_K}$, the ℓth order polynomial spline is the function $x(t)$ which obeys:

$$x(t_k) = v_{t_k}, \text{ for } k = 1, \ldots, K,$$

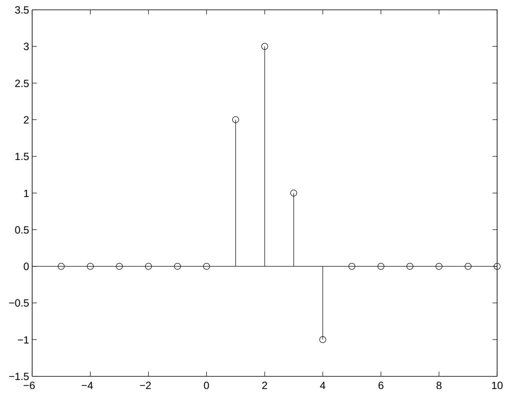
and

$$x(t) \text{ is an } \ell\text{th order polynomial between the } t_k$$

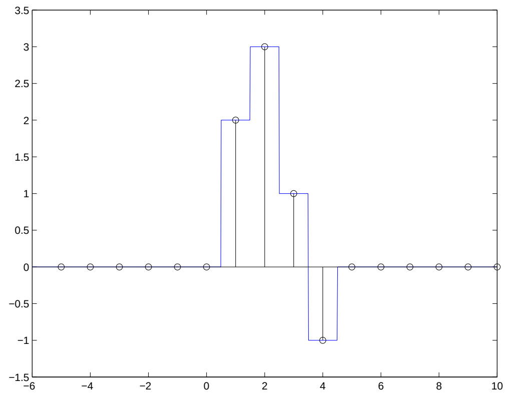
For $\ell \geq 1$, the spline function is continuous and will have $\ell - 1$ derivatives which are continuous at the t_k.

For example, if we have data points at the integers
$t_1 = 1,\ t_2 = 2,\ t_3 = 3,\ t_4 = 4$

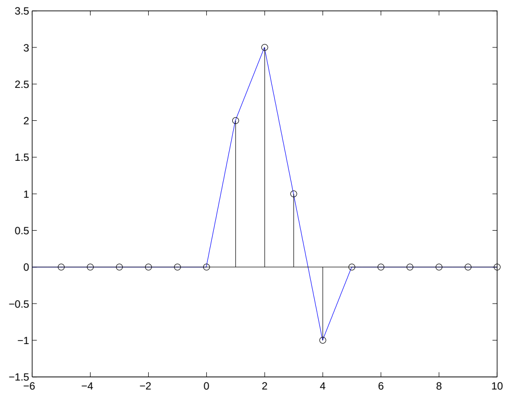
with values
$v_1 = 2,\ v_2 = 3,\ v_3 = 1,\ v_4 = -1$ and values of zeros at the other integers,



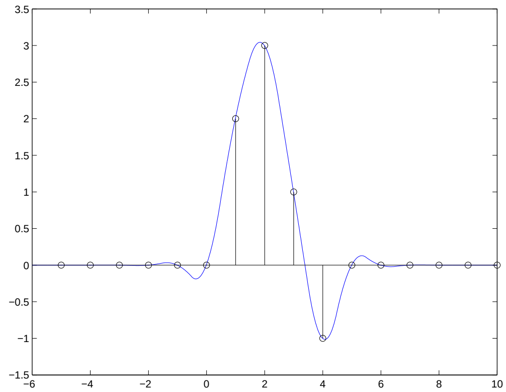
then here is the zero-th order interpolation,



the linear interpolation,



and the quadratic interpolation,



Any ℓth order polynomial spline can be written as a superposition of B-spline functions (the 'B' is for basis!).

The piecewise constant function above can be written as

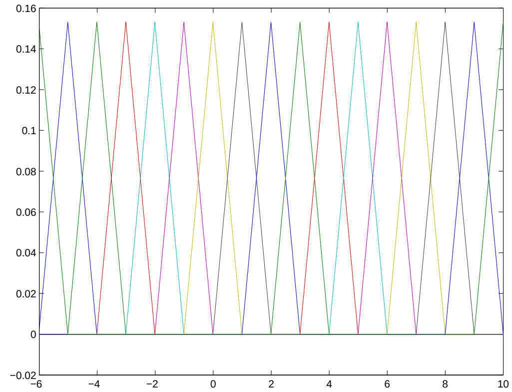
$$x(t) = \sum_{k=1}^{4} \alpha_k \, b_0(t-k), \quad b_0(t) = \begin{cases} 1 & -1/2 \leq t < 1/2 \\ 0 & \text{else} \end{cases}$$

for $\alpha_1 = 2,\ \alpha_2 = 3,\ \alpha_3 = 1,\ \alpha_4 = -1.$ The piecewise linear function above can be written as

$$x(t) = \sum_{k=1}^{4} \alpha_k \, b_1(t-k), \quad b_1(t) = \begin{cases} t+1 & -1 \leq t \leq 0 \\ 1-t & 0 \leq t \leq 1 \\ 0 & \text{else} \end{cases}$$

for $\alpha_1 = 2,\ \alpha_2 = 3,\ \alpha_3 = 1,\ \alpha_4 = -1.$ In this case, the building blocks $b_1(t)$ are 'hat' functions:

$b_1(t-k),\ k=-6,\ldots,10$



For spline expansions using an order greater than 1, the expansion coefficients α_k will not be equal to the sample values. However, given a set of M samples values, the α_k that interpolate these samples can be found by solving a system of equations.

Notice that we can generate b_1 by convolving b_0 with itself:

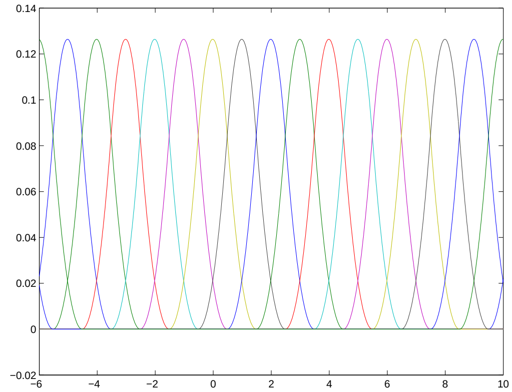
$$b_1(t) = (b_0 * b_0)(t).$$

The expansion for the piecewise quadratic spline above is a little more complicated:

$$x(t) = \sum_{k=-\infty}^{\infty} \alpha_k\, b_2(t-k), \quad b_2(t) = \begin{cases} (t+3/2)^2/2 & -3/2 \le t \le -1/2 \\ -t^2 + 3/4 & -1/2 \le t \le 1/2 \\ (t-3/2)^2/2 & 1/2 \le t \le 3/2 \\ 0 & |t| \ge 3/2 \end{cases}$$

where the α_k are all non-zero. The expansion has an infinite number of terms because the basis functions overlap on the integers:

$b_2(t-k),\ k=-6,\ldots,10$



Just as before, we can generate the basis function $b_2(t)$ from the lower order ones:

$$b_2(t) = (b_1 * b_0)(t) = (b_0 * b_0 * b_0)(t).$$

In general, any ℓth order polynomial spline $x(t)$ is uniquely represented by a list of numbers $\{\alpha_k,\ k \in \mathbb{Z}\}$, which correspond to the weights needed to re-synthesize the spline from the building blocks $\{b_\ell(t-k),\ k \in \mathbb{Z}\}$:

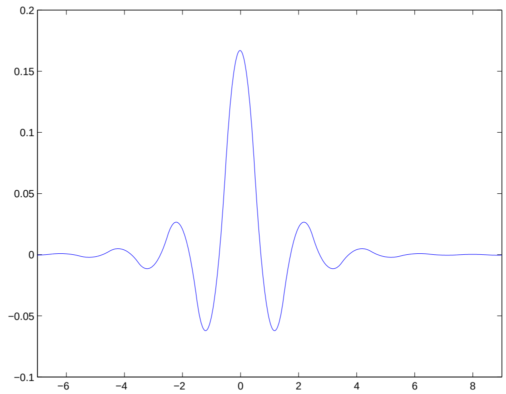
$$x(t) = \sum_{k=-\infty}^{\infty} \alpha_k b_\ell(t-k), \quad b_\ell(t) = \underbrace{b_0(t) * \cdots * b_0(t)}_{\ell \text{ times}}$$

If we are given an ℓth order spline $x(t)$, there is a systematic way to compute the corresponding α_k (and hence gives us another reproducing formula). Without getting too much into the details at this point, there is a complementary function $\tilde{b}_\ell(t)$ such that

$$\alpha_k = \int_{-\infty}^{\infty} x(t)\tilde{b}_\ell(t-k)\ \mathrm{d}t.$$

In the quadratic case $\ell = 2$, this function looks kind of like a sinc:

$\tilde{b}_2(t)$



We know how to compute these complementary $\tilde{b}_\ell(t)$, but they are not easy to write down as nice expressions.

Bases and discretization

All of the examples above have a common theme: we take a signal in a certain class (bandlimited, zero outside of $[0, T]$, polynomial spline) and represent it using a discrete list of numbers $\{\alpha_k,\ k \in \mathbb{Z}\}$.

These numbers represent weights used to build up the signal out of a set of pre-determined building blocks ("basis functions"). This framework gives us a systematic way to manipulate continuous time signals by operating on discrete vectors. This allows us to unleash the power of **linear algebra**.

It often times also gives us a straightforward way to simply or compress signals. As you can see from the sawtooth Fourier series example, although it technically takes an infinity of sinusoids to build up the signals, we can get away with 50 if we are willing to suffer some loss. We will see some more examples of this later in this section.

In this set of notes, we have just gotten our first taste of basis expansions. What we will do next is develop a systematic method for taking a signal and breaking it down into a superposition of basis functions. We will also discuss how to optimally approximate a function using a fixed number of basis functions — this simple idea has an incredible number of applications.

To do these things correctly, we need to first build up some mathematical machinery so we can avoid talking in hazy terms. We start in the next section with precise (but abstract) definitions of **linear vector space**, **norm**, and **inner product**.

Exercises

1. Consider the set of signals that can be written as

$$x(t) = \alpha_0 b_2(t) + \alpha_1 b_2(t-1).$$

Suppose you know that $x(0) = 1$ and $x(1) = -1$. What must α_0 and α_1 be?

2. Consider the set of signals that can be written as

$$x(t) = \alpha_0 b_0(t) + \alpha_1 b_0(t-1) + \alpha_2 b_0(t-2) + \alpha_3 b_0(t-3).$$

How to I find the $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ such that $x(t)$ is as close[2] to $\cos(t)$ as possible?

[2]Of course, the answer will depend on how you define "close"...