An Introduction to Orthogonal Polynomials

by Marek Rychlik

March 14, 2009

Orthogonal polynomials in function spaces

We tend to think of scientific data as having some sort of continuity. This allows us to approximate these data by special functions, such as polynomials or finite trigonometric series. The quantitative measure of the quality of these approximations is necessary. It is typically given by a norm.

Definition 1. *The space of square-integrable functions on the interval* $[-1,1]$ *is a vector space consisting of all measurable functions* $f\colon [-1,1] \to \mathbb{R}$ *such that*

$$\int_1^1 f(x)^2 dx < \infty.$$

The integral is in the sense of Lebesgue.

This space is denoted by $L^2([-1,1])$. The space is endowed with an inner product:

$$\langle f, g \rangle = \int_{-1}^1 f(x) g(x) dx.$$

Remarks on convergence

Definition 2. *A Cauchy sequence in a metric space* (V, d) *where* $d: V \times V \to \mathbb{R}$ *is a metric, is a sequence* $(x_n)_{n=1}^{\infty}$ *such that* $x_n \in V$ *and for every* $\epsilon > 0$ *there is* N *such that for all* $m, n \geq N$ *we have:*

$$d(x_m, x_n) < \epsilon.$$

Definition 3. *A Banach space is a normed space* $(V, \|\cdot\|)$ *which is complete as a metric space, i.e. in which every Cauchy sequence converges. The metric is given by* $d(u, v) = \|u - v\|$.

Definition 4. *A Hilbert space is an inner product space* $(V, \langle\,\cdot\,,\,\cdot\,\rangle)$ *which is a Banach space as a normed space with the norm* $\|u\| = \sqrt{\langle u, u\rangle}$.

Theorem 5. $L^2([-1, 1])$ *is a Hilbert space.*

Orthogonal sets in L^2([-1,1])

Studying orthogonality in $L^2([-1,1])$ has been one of the most fruitful human endeavors, as it let to the advent of Fourier Theory and its modern continuation, the wavelet theory.

It is a standard result in Fourier series theory, that the following set is orthonormal:

$$\{1\} \cup \{\cos(n\pi x)\}_{n=1}^{\infty} \cup \{\sin(n\pi x)\}_{n=1}^{\infty}$$

This is equivalent to the vanishing of certain integrals of trigonometric functions:

$$\begin{aligned}
\int_{-1}^{1} \cos(n\pi x)\,dx &= 0, \\
\int_{-1}^{1} \cos(n\pi x)\cos(m\pi x)\,dx &= 0 \text{ for } m \neq n, \\
\int_{-1}^{1} \cos(n\pi x)\sin(n\pi x)\,dx &= 0, \\
\int_{-1}^{1} \sin(n\pi x)\sin(m\pi x)\,dx &= 0 \text{ for } m \neq n, \\
\int_{-1}^{1} \cos^2(n\pi x)\,dx &= 1,
\end{aligned}$$

$$\int_{-1}^{1} \sin^2(n\pi x)\,dx \;=\; 1.$$

Theorem 6. *Every function $f \in L^2([-1,1])$ admits a Fourier series representation:*

$$f = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos(n\pi x) + b_n \sin(n\pi x)\right)$$

where:

$$a_n \;=\; \langle f, \cos(n\pi x)\rangle = \int_{-1}^{1} f(x)\cos(n\pi x)\,dx \quad \text{for } n = 0, 1, \ldots,$$

$$b_n \;=\; \langle f, \sin(n\pi x)\rangle = \int_{-1}^{1} f(x)\sin(n\pi x)\,dx \quad \text{for } n = 1, 2, \ldots.$$

The equality in the representation means that:

$$\lim_{N\to\infty} \left\| f - \left(\frac{a_0}{2} + \sum_{n=1}^{N} \left(a_n \cos(n\pi x) + b_n \sin(n\pi x)\right)\right) \right\| = 0.$$

It does not mean pointwise convergence of the right-hand side to the value of $f(x)$.

Hilbert bases

Definition 7. *A Hilbert basis is an orthogonal subset* $\{e_1, e_2, ...\}$ *in a Hilbert space* $(V, \langle \cdot, \cdot \rangle)$ *such that for every* $f \in V$ *there and every* $\epsilon > 0$ *there exists a sequence of numbers* α_i*, a finite number of which is* $\neq 0$ *and such that:*

$$\| f - \sum_{i=1}^{\infty} \alpha_i e_i \| < \epsilon.$$

Remark 8. Thus, we assume that every element of V can be approximated by finite linear combinations of the elements of the orthogonal set.

Theorem 9. *If* $\{e_1, e_2, ...\}$ *is a Hilbert basis in a Hilbert space* $(V, \langle \cdot, \cdot \rangle)$ *then:*

$$\lim_{N \to \infty} \left\| f - \sum_{i=1}^{N} \frac{\langle f, e_i \rangle}{\langle e_i, e_i \rangle} e_i \right\| = 0.$$

In orther words, the sequence of projections of f onto $\text{span}\{e_1, e_2, \ldots, e_N\}$ *converges to f as $N \to \infty$.*

Legendre polynomials

In many applications, polynomials are preferred to trigonometric functions, for many reasons, e.g. the cost of numerical evaluation.

We have already examined the Gram-Schmidt process for converting any linearly independent set to an orthogonal set. We may apply Gram-Schmidt process to the sequence of powers $\{1, x, x^2, \ldots\}$ to obtain an infinite orthogonal set. The polynomials we obtain are:

$$
\begin{aligned}
Q_0(x) &= 1, \\
Q_1(x) &= x - \frac{\langle x, 1\rangle}{\langle 1, 1\rangle} 1 = x, \text{because } \langle x, 1\rangle = 0, \\
Q_2(x) &= x^2 - \frac{\langle x^2, 1\rangle}{\langle 1, 1\rangle} 1 - \frac{\langle x^2, x\rangle}{\langle x, x\rangle} x = x^2 - \frac{1}{3}.
\end{aligned}
$$

In similar fashion, we can obtain additional Legendre polynomials. The theory of Legendre polynomials yields the following expression (the Rodrigues formula):

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$$

which is equivalent to ours, up to the normalizing constant. We can see that each of our polynomials $Q_n(x)$ has coefficient 1 at power n. Rodrigues formula yields the coefficient at x^n euqal to:

$$\frac{(2n)(2n-1)...(n+1)}{2^n n!} = \frac{(2n)!}{2^n (n!)^2} = \frac{1}{2^n}\binom{2n}{n}.$$

Hence, the formula:

$$P_n(x) = \frac{1}{2^n}\binom{2n}{n} Q_n(x).$$

The Legendre polynomials are orthogonal, and their normalizing constants are obtained from the formula:

$$\langle P_n, P_n\rangle = \int_{-1}^{1} P_m(x)^2 dx = \frac{2}{2n+1}.$$

Computing first few Legendre polynomials

We use an open source Computer Algebra System (CAS) called Maxima, to compute the first few Legendre polynomials:

```
(%i1) Q[0](x):= 1;
```

(%o16) $Q_0(x) := 1$

```
(%i17) Q[n](x):= expand(x^n-sum(integrate(x^n*Q[k](x), x, -1,
       1)/integrate(Q[k](x)*Q[k](x),x,-1,1)*Q[k](x), k, 0, n-1));
```

(%o17) $Q_n(x) := \mathrm{expand}\left(x^n - \mathrm{sum}\left(\frac{\mathrm{integrate}(x^n\, Q_k(x), x, -1, 1)}{\mathrm{integrate}(Q_k(x)\, Q_k(x), x, -1, 1)}\, Q_k(x), k, 0, n-1\right)\right)$

```
(%i18) for k from 0 thru 7 do ( display(Q[k](x)) );
```

$Q_0(x) = 1$

$Q_1(x) = x$

$$Q_2(x) = x^2 - \frac{1}{3}$$

$$Q_3(x) = x^3 - \frac{3\,x}{5}$$

$$Q_4(x) = x^4 - \frac{6\,x^2}{7} + \frac{3}{35}$$

$$Q_5(x) = x^5 - \frac{10\,x^3}{9} + \frac{5\,x}{21}$$

$$Q_6(x) = x^6 - \frac{15\,x^4}{11} + \frac{5\,x^2}{11} - \frac{5}{231}$$

$$Q_7(x) = x^7 - \frac{21\,x^5}{13} + \frac{105\,x^3}{143} - \frac{35\,x}{429}$$

(%o18) **done**

```
(%i19) plot2d(makelist(Q[k](x),k,0,7),[x,-
       1,1],[psfile,"/tmp/Q.ps"]);
```

(%o19)

```
(%i20)
```

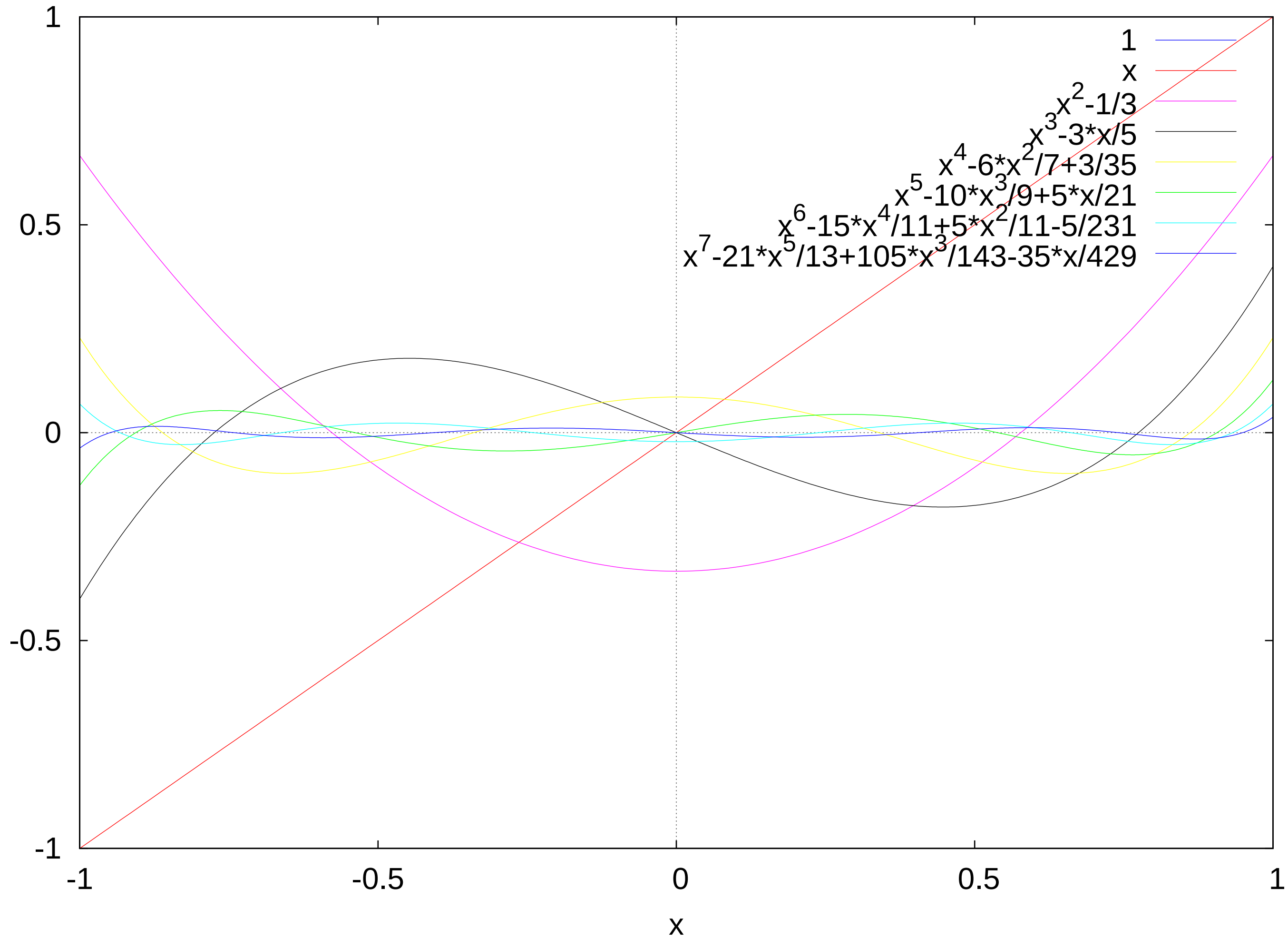

Figure 1. The plot of the first 8 Q_k's.

If we use Rodrigues' formula, we obtain a slightly different plot. Here is the calculation:

```
(%i12) P[n](x):=expand((1/(2^n*n!))*diff((x^2-1)^n,x,n));
```

(%o9) $$P_n(x):=\mathrm{expand}\left(\frac{1}{2^n\,n!}\,\mathrm{diff}\left(\left(x^2-1\right)^n,x,n\right)\right)$$

```
(%i10) for k from 0 thru 5 do ( display(P[k](x)) );
```

$$P_0(x) = 1$$

$$P_1(x) = x$$

$$P_2(x) = \frac{3\,x^2}{2} - \frac{1}{2}$$

$$P_3(x) = \frac{5\,x^3}{2} - \frac{3\,x}{2}$$

$$P_4(x) = \frac{35\,x^4}{8} - \frac{15\,x^2}{4} + \frac{3}{8}$$

$$P_5(x) = \frac{63\,x^5}{8} - \frac{35\,x^3}{4} + \frac{15\,x}{8}$$

(%o10) **done**

```
(%i11) plot2d(makelist(P[k](x),k,0,7),[x,-
       1,1],[psfile,"/tmp/P.ps"])$
```

```
(%i21)
```

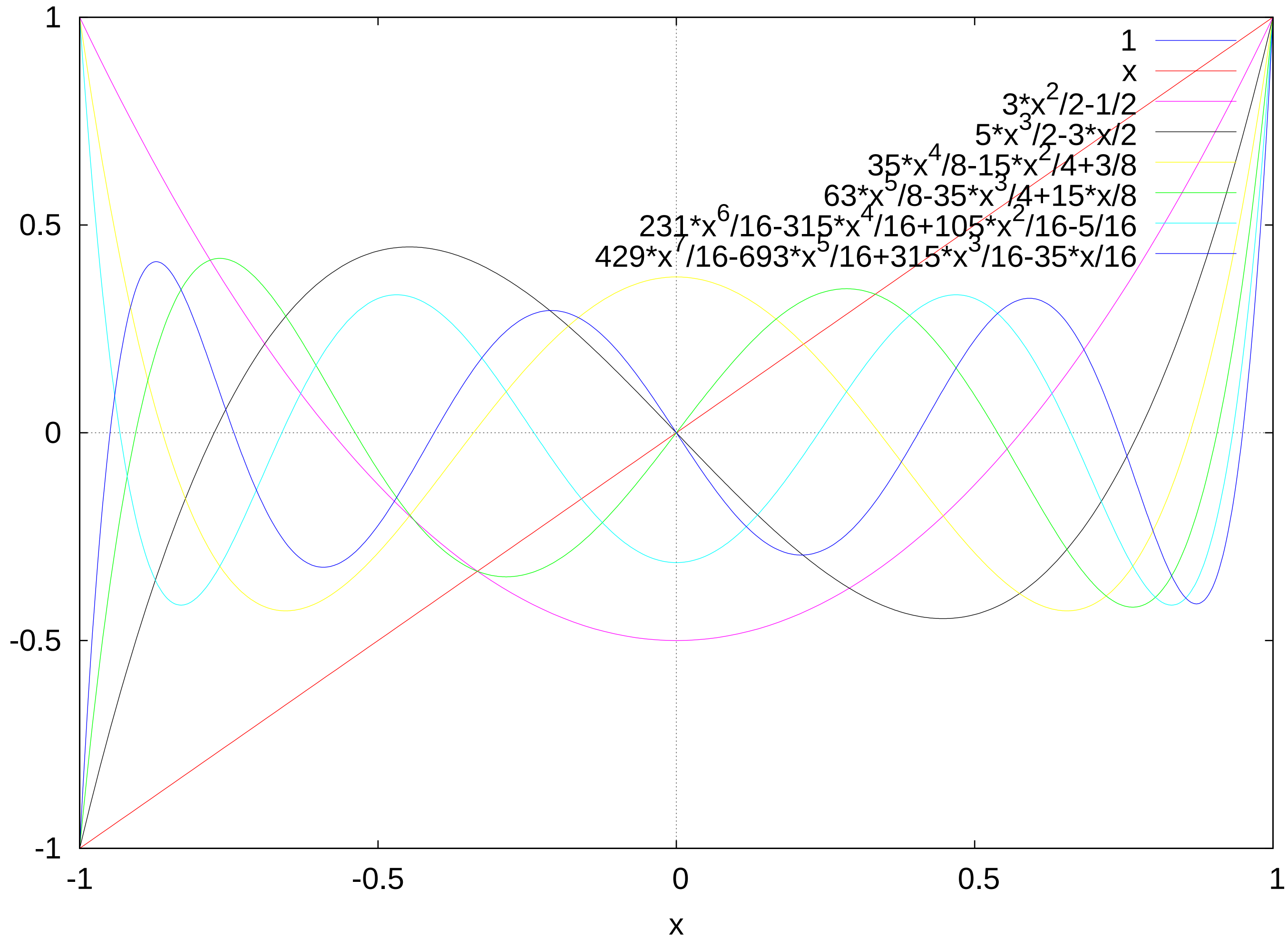

Figure 2. The plot of the first 8 P_k's.

We can see that P_k's are scaled so that the value at 1 is $+1$.

Orthogonal polynomials in Statistics

The polynomials commonly used as orthogonal contrasts for quantitative factors are discrtete analogues of Legendre polynomials. One way to understand them is to consider the discretization of the inner product of $L^2([a,b])$:

$$\langle f, g \rangle = \sum_{i=0}^{t-1} f(x_i) g(x_i)$$

where x_i is an increasing sequence of points in $[a, b]$. The most common case is that of equally spaced points:

$$x_i = a + i \cdot d$$

where $d = \frac{b-a}{t}$. We may replace the variable x (the factor) with i by performing the mapping:

$$x \mapsto \tilde{x} = \frac{x - \bar{x}}{d}$$

where $\bar{x}$ is the mean of x_i. We can see that the values assumed by the transformed variable are $\tilde{x}_i = i - \frac{t-1}{2}$, $i = 0, 1, \ldots, t-1$.

Because we are using only a finite number of points, the bilinear form just defined is degenerate, i.e. it is possible that $\langle f, g \rangle = 0$ for all g and still $f \neq 0$. However, if we restrict this form to the set of polynomials of degree t, the form becomes non-degenerate. Hence, we consider the vector space V_t of all polynomials of degree $< t$:

$$f(x) = \sum_{i=0}^{t-1} \beta_i x^i.$$

Lemma 10. *The space $(V_t, \langle \cdot, \cdot \rangle)$ is an inner product space.*

Proof. We need to study the Van der Monde matrix. □

Van der Monde matrix

Definition 11. *Van der Monde matrix is defined for a sequence of points $(x_1, x_2, \ldots, x_t)$ as follows:*

$$M = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{t-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{t-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_t & x_t^2 & \ldots & x_t^{t-1} \end{pmatrix}.$$

Lemma 12. *The determinant of M is:*

$$\det(M) = \prod_{i<j} (x_i - x_j).$$

In particular, if x_i are all distinct then $\det(M) \neq 0$.

Lemma 13. *The mapping $F: V_t \to \mathbb{R}^t$ which maps a polynomial to the vector of its values:*

$$f(x) \mapsto \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_t) \end{pmatrix}$$

has the Van der Monde matrix as its matrix, if the basis $\{1, x, ..., x^{t-1}\}$ is used as a basis of V_t and the standard basis is used as a basis of $\mathbb{R}^t$. If x_i are all distinct then F is non-singular, and thus an isomorphism.

Lagrange interpolating polynomials

The problem of finding a polynomial $f \in V_t$ which assumes given values $(b_1, b_2, ..., b_t)$ at given points x_1, x_2, ..., x_t is the Lagrange interpolation problem and it is equivalent to finding the inverse of F. The solutions takes the form of the

Lagrange interpolating polynomial. There are many ways to derive the formula for the Lagrange interpolating polynomial, and one of them is to use Cramer's Rule to solve the linear system $M\beta = b$, which is obtained by rewriting the equation $F(f) = b$ in the aforementioned bases of V_t and $\mathbb{R}^t$. The solution is:

$$
\begin{aligned}
f(x) &= \sum_{i=1}^{t} b_i L_{t,i}(x), \quad \text{where} \\
L_{t,i}(x) &= \frac{\displaystyle\prod_{\substack{j=1,2,\ldots,t \\ j\neq i}} (x - x_j)}{\displaystyle\prod_{\substack{j=1,2,\ldots,t \\ j\neq i}} (x_i - x_j)} \quad \text{for } i = 1, 2, \ldots, t.
\end{aligned}
$$

The polynomials $L_{t,i}(x)$ are the unique polynomials of degree $t-1$ such that

$$L_{t,i}(x_j) = \delta_{ij}$$

where δ_{ij} is the Kronecker delta.

Gram-Schmidt orthogonalization in V_t

We perform the Gram-Schmidt process on the basis $\{1, x, \ldots, x^{t-1}\}$ in V_t.

We assume that we transformed the variable, so that $x_i = i - (t-1)/2$ for $i = 0$, $1, ..., t-1$. Here are the first few polynomials computed by hand:

$$\begin{aligned}
P_0(x) &= 1, \\
P_1(x) &= x - \frac{\langle x, 1\rangle}{\langle 1, 1\rangle} 1 = x, \\
P_2(x) &= x^2 - \frac{\langle x^2, 1\rangle}{\langle 1, 1\rangle} 1 - \frac{\langle x^2, x\rangle}{\langle x, x\rangle} x \\
&= x^2 - \frac{\sum_{i=0}^{t-1} x_i^2}{\sum_{i=0}^{t-1} 1} 1 - \frac{\sum_{i=0}^{t-1} x_i^3}{\sum_{i=0}^{t-1} x_i^2} x \\
&= x^2 - \frac{t^2-1}{12}.
\end{aligned}$$

We can see that in the process of calculations, we need to use closed-form formulas for the sums of the powers of integers:

$$b_k(t) = \sum_{i=0}^{t-1} x_i^k.$$

The theory leading to these closed form formulas is well know and has to do with Bernoulli polynomials. We will only note that by induction it is easy to prove that $b_k(t)$ is a polynomial of degree $k+1$ in t. We can use Maxima to produce closed formulas for the sums used in the above calculation. We produce $b_k(t)$ for even t only, as for odd k $b_k(t)=0$ for reasons of parity.

```
(%i77) b[k](t):=nusum((i-(t-1)/2)^k,i,0,t-1);
```

(%o86) $$b_k(t) := \mathrm{nusum}\left(\left(i-\frac{t-1}{2}\right)^k, i, 0, t-1\right)$$

```
(%i87) for k from 0 thru 6 step 2 do ( display(b[k](t)) );
```

$$b_0(t) = t$$

$$b_2(t) = \frac{(t-1)\,t\,(t+1)}{12}$$

$$b_4(t) = \frac{(t-1)\,t\,(t+1)\left(3\,t^2-7\right)}{240}$$

$$b_6(t) = \frac{(t-1)\,t\,(t+1)\left(3\,t^4-18\,t^2+31\right)}{1344}$$

(%o87) **done**

```
(%i88)
```

Computing orthogonal polynomials with a CAS

We once again employ maxima to compute the orthogonal polynomials used as contrasts in statistics. For simplicity, we fix the order at the beginning.

```
(%i88) kill(t,offset,p,inner,norm,B)
```

(%o115) **done**

```
(%i116) t:7;
```

(%o120) 7

```
(%i121) offset:nusum(i,i,0,t-1)/t;
```

(%o121) $\frac{37}{7}$

```
(%i122) p[i]:=i-offset;
```

(%o122) $p_i := i - \mathrm{offset}$

```
(%i123) inner(f,g):=sum(f(p[i])*g(p[i]),i,0,t-1);
```

(%o123) $\mathrm{inner}(f, g) := \mathrm{sum}(f(p_i)\, g(p_i), i, 0, t-1)$

```
(%i124) norm(f):=sqrt(inner(f,f));
```

(%o125) $\mathrm{norm}(f) := \sqrt{\mathrm{inner}(f, f)}$

```
(%i126) B[0](x):=1;
```

(%o126) $B_0(x) := 1$

```
(%i127) B[n](x):=expand(x^n-
        sum(inner(lambda([x],x^n),B[k])/inner(B[k],B[k])
        *B[k](x),k,0,n-1));
```

(%o127) $B_n(x) := \mathrm{expand}\left(x^n - \mathrm{sum}\left(\frac{\mathrm{inner}(\lambda([x], x^n), B_k)}{\mathrm{inner}(B_k, B_k)}\, B_k(x), k, 0, n-1\right)\right)$

```
(%i128) for k from 0 thru t-1 do( display(B[k](x)) )$
```

$$B_0(x) = 1$$

$$B_1(x) = x$$

$$B_2(x) = x^2 - 4$$

$$B_3(x) = x^3 - 7\,x$$

$$B_4(x) = x^4 - \frac{67\,x^2}{7} + \frac{72}{7}$$

$$B_5(x) = x^5 - \frac{35\,x^3}{3} + \frac{524\,x}{21}$$

$$B_6(x) = x^6 - \frac{145\,x^4}{11} + \frac{434\,x^2}{11} - \frac{1200}{77}$$

```
(%i129) plot2d(makelist(B[k](x),k,0,t-1),
         [x,p[0],p[t-1]],
         [psfile,"/tmp/B.ps"])$
```

```
(%i132)
```

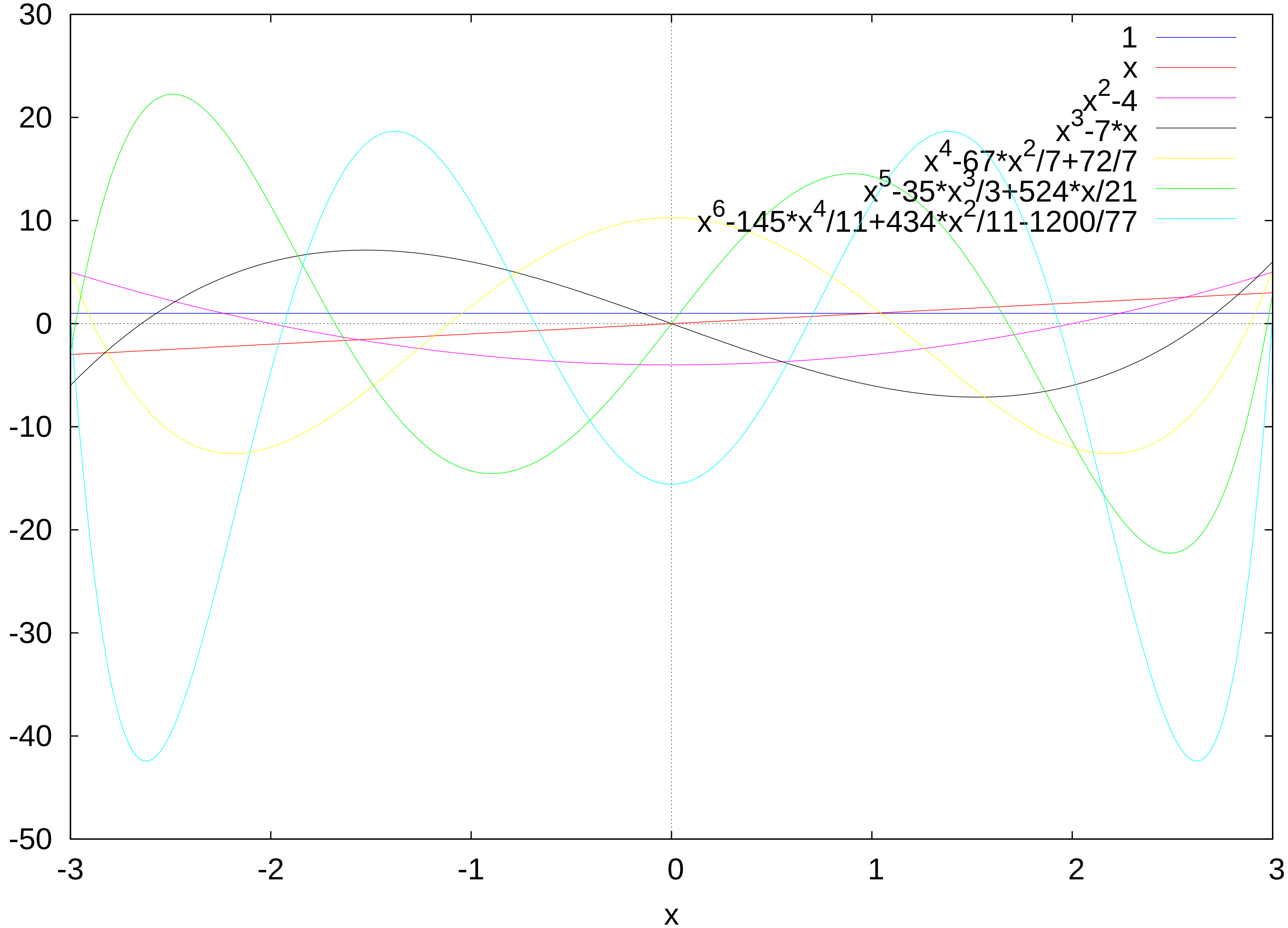

Figure 3. A plot of the B_k's for $t = 7$.