July 20, 2013

Chapter 12: Best Approximation

Uri M. Ascher and Chen Greif
Department of Computer Science
The University of British Columbia
{ascher,greif}@cs.ubc.ca

Slides for the book
A First Course in Numerical Methods (published by SIAM, 2011)
`http://www.ec-securehost.com/SIAM/CS07.html`

Goals of this chapter

- To develop some elegant, classical theory of best approximation using least squares and weighted least squares norms;
- to develop important families of orthogonal polynomials;
- to examine in particular Legendre and Chebyshev polynomials.

Outline

- Best least squares approximation
- Orthogonal basis functions
- Weighted least squares
- Chebyshev polynomials

Approximation, not just interpolation

- As in Chapters 10 and 11, we are given a complicated (or implicit) function $f(x)$ which may be evaluated anywhere on the interval $[a, b]$, and we look for a simpler approximation $v(x) = \sum_{j=0}^{n} c_j \phi_j(x)$, where $\phi_j(x)$ are known basis functions with some useful properties.
- Unlike Chapters 10 and 11, we no longer seek to impose $v(x_i) - f(x_i) = 0$ at any $n + 1$ specific interpolation points.
- One reason for shunning interpolation: if there is significant noise in the measured values of f then measuring the approximation error at $m + 1 > n + 1$ points may yield more plausible or meaningful results. This leads to **overdetermined** linear systems which are most easily solved using discrete least squares methods; see **Section 6.1**.
- Here, however, we stick to the continuous, rather than discrete level, and use integrals rather than vectors. Typically we may seek a function $v(x)$, i.e., coefficients c_j, so as to minimize

$$\|v - f\|_2^2 = \int_a^b (v(x) - f(x))^2 dx.$$

Aside: function norms

- For all integrable functions $g(x)$ and $f(x)$ on an interval $[a,b]$, a norm $\|\cdot\|$ is a scalar function satisfying
  - $\|g\| \geq 0$; $\|g\| = 0$ iff $g(x) \equiv 0$,
  - $\|\alpha g\| = |\alpha|\|g\|$ for all scalars α,
  - $\|f+g\| \leq \|f\| + \|g\|$.

  The set of all functions whose norm is finite forms a function space associated with that particular norm.
- Some popular function norms and corresponding function spaces are:

$$\begin{aligned} L_2: \quad \|g\|_2 &= \left(\int_a^b g(x)^2 dx\right)^{1/2}, \quad \text{(least squares)}, \\ L_1: \quad \|g\|_1 &= \int_a^b |g(x)| dx, \\ L_\infty: \quad \|g\|_\infty &= \max_{a \leq x \leq b} |g(x)|, \quad \text{(max)}. \end{aligned}$$

- We have seen the L_∞ norm in action when estimating errors in Chapters 10 and 11. Here we concentrate on the L_2 norm, occasionally in weighted form.

Aside: function norms

- For all integrable functions $g(x)$ and $f(x)$ on an interval $[a, b]$, a norm $\|\cdot\|$ is a scalar function satisfying
  - $\|g\| \geq 0$; $\|g\| = 0$ iff $g(x) \equiv 0$,
  - $\|\alpha g\| = |\alpha|\|g\|$ for all scalars α,
  - $\|f + g\| \leq \|f\| + \|g\|$.

  The set of all functions whose norm is finite forms a function space associated with that particular norm.
- Some popular function norms and corresponding function spaces are:

$$
\begin{aligned}
L_2: \quad \|g\|_2 &= \left(\int_a^b g(x)^2 dx\right)^{1/2}, \quad \text{(least squares)},\\
L_1: \quad \|g\|_1 &= \int_a^b |g(x)| dx,\\
L_\infty: \quad \|g\|_\infty &= \max_{a \leq x \leq b} |g(x)|, \quad \text{(max)}.
\end{aligned}
$$

- We have seen the L_∞ norm in action when estimating errors in Chapters 10 and 11. Here we concentrate on the L_2 norm, occasionally in weighted form.

Aside: function norms

- For all integrable functions $g(x)$ and $f(x)$ on an interval $[a, b]$, a norm $\| \cdot \|$ is a scalar function satisfying
    - $\|g\| \geq 0$; $\|g\| = 0$ iff $g(x) \equiv 0$,
    - $\|\alpha g\| = |\alpha| \|g\|$ for all scalars α,
    - $\|f + g\| \leq \|f\| + \|g\|$.

  The set of all functions whose norm is finite forms a function space associated with that particular norm.
- Some popular function norms and corresponding function spaces are:

$$
\begin{aligned}
L_2: \quad \|g\|_2 &= \left(\int_a^b g(x)^2 dx \right)^{1/2}, \quad \text{(least squares)}, \\
L_1: \quad \|g\|_1 &= \int_a^b |g(x)| dx, \\
L_\infty: \quad \|g\|_\infty &= \max_{a \leq x \leq b} |g(x)|, \quad \text{(max)}.
\end{aligned}
$$

- We have seen the L_∞ norm in action when estimating errors in Chapters 10 and 11. Here we concentrate on the L_2 norm, occasionally in weighted form.

Why continuous approximation?

- Discrete least squares (Chapter 6) is certainly efficiently manageable in many situations and readily applicable. Indeed, in this chapter we don't directly concentrate on efficient programs or algorithms.
- However, sticking to the continuous allows developing a mathematically elegant theory.
- Moreover, classical families of orthogonal basis functions are developed which prove highly useful in Chapters 13, 15, 16 and others.

Outline

- Best least squares approximation
- Orthogonal basis functions
- Weighted least squares
- Chebyshev polynomials

Best least squares approximation

- We are still approximating a given $f(x)$ on interval $[a, b]$ using

$$v(x) = \sum_{j=0}^{n} c_j \phi_j(x),$$

but we are no longer necessarily interpolating.

- Determine the coefficients c_j as those that solve the minimization problem

$$\min_{\mathbf{c}} \|f - v\|^2 \equiv \min_{\mathbf{c}} \int_a^b \left[f(x) - \sum_{j=0}^{n} c_j \phi_j(x) \right]^2 dx.$$

- Taking first derivative with respect to each c_k in turn and equating to 0 gives

$$\int_a^b \left[f(x) - \sum_{j=0}^{n} c_j \phi_j(x) \right] \phi_k(x) dx = 0, \quad k = 0, 1, \ldots, n.$$

Best least squares approximation

- We are still approximating a given $f(x)$ on interval $[a, b]$ using

$$v(x) = \sum_{j=0}^{n} c_j \phi_j(x),$$

but we are no longer necessarily interpolating.

- Determine the coefficients c_j as those that solve the minimization problem

$$\min_{\mathbf{c}} \|f - v\|^2 \equiv \min_{\mathbf{c}} \int_a^b \left[f(x) - \sum_{j=0}^{n} c_j \phi_j(x) \right]^2 dx.$$

- Taking first derivative with respect to each c_k in turn and equating to 0 gives

$$\int_a^b \left[f(x) - \sum_{j=0}^{n} c_j \phi_j(x) \right] \phi_k(x) dx = 0, \quad k = 0, 1, \ldots, n.$$

Best least squares approximation cont.

- This yields the following **construction algorithm**:
  1. Calculate
$$\tilde{B}_{j,k} = \int_a^b \phi_j(x)\phi_k(x)dx, \quad j,k = 0,1,\ldots,n$$
$$\tilde{b}_j = \int_a^b f(x)\phi_j(x)dx \quad j = 0,1,\ldots,n.$$
  2. Solve linear $(n+1)\times(n+1)$ system
$$\tilde{B}\mathbf{c} = \tilde{\mathbf{b}}$$
for the coefficients $\mathbf{c} = (c_0, c_1, \ldots, c_n)^T$.
- **Example**: using a monomial basis $\phi_j(x) = x^j$ for a polynomial approximation on $[0,1]$, obtain
$$\tilde{B}_{j,k} = \int_0^1 x^{j+k}dx = 1/(j+k+1), \quad 0 \le j,k \le n.$$
Also, $\tilde{b}_j = \int_0^1 f(x)x^j dx$ can be approximated by Chapter 15 techniques.

Best least squares approximation cont.

- This yields the following **construction algorithm**:
  1. Calculate
  $$\tilde{B}_{j,k} = \int_a^b \phi_j(x)\phi_k(x)dx, \quad j,k = 0,1,\ldots,n$$
  $$\tilde{b}_j = \int_a^b f(x)\phi_j(x)dx \quad j = 0,1,\ldots,n.$$
  2. Solve linear $(n+1) \times (n+1)$ system
  $$\tilde{B}\mathbf{c} = \tilde{\mathbf{b}}$$
  for the coefficients $\mathbf{c} = (c_0, c_1, \ldots, c_n)^T$.
- **Example**: using a monomial basis $\phi_j(x) = x^j$ for a polynomial approximation on $[0,1]$, obtain
$$\tilde{B}_{j,k} = \int_0^1 x^{j+k}dx = 1/(j+k+1), \quad 0 \le j,k \le n.$$
Also, $\tilde{b}_j = \int_0^1 f(x)x^j\,dx$ can be approximated by Chapter 15 techniques.

Outline

- Best least squares approximation
- Orthogonal basis functions
- Weighted least squares
- Chebyshev polynomials

Orthogonal basis functions

- The simple example using monomials leads to stability problems (bad conditioning in $\tilde{B}$).
- Using monomial and other simple-minded bases, must evaluate many integrals when n is large.
- Idea: construct basis functions such that $\tilde{B}$ is diagonal!
  This means

$$\int_a^b \phi_j(x)\phi_k(x)dx = 0, \quad j \neq k.$$

- Take advantage of the fact that here, unlike interpolation, there are no input abscissae: can construct families of orthogonal basis functions in advance!

Orthogonal basis functions

- The simple example using monomials leads to stability problems (bad conditioning in $\tilde{B}$).
- Using monomial and other simple-minded bases, must evaluate many integrals when n is large.
- Idea: construct basis functions such that $\tilde{B}$ is diagonal!
  This means

$$\int_a^b \phi_j(x)\phi_k(x)dx = 0, \quad j \neq k.$$

- Take advantage of the fact that here, unlike interpolation, there are no input abscissae: can construct families of orthogonal basis functions in advance!

Orthogonal basis functions

- The simple example using monomials leads to stability problems (bad conditioning in $\tilde{B}$).
- Using monomial and other simple-minded bases, must evaluate many integrals when n is large.
- Idea: construct basis functions such that $\tilde{B}$ is diagonal!
  This means

$$\int_a^b \phi_j(x)\phi_k(x)dx = 0, \quad j \neq k.$$

- Take advantage of the fact that here, unlike interpolation, there are no input abscissae: can construct families of orthogonal basis functions in advance!

Best least squares using orthogonal basis functions

- The **construction algorithm** becomes simply:
  1. For $j = 0, 1, \ldots, n$, set $d_j = \tilde{B}_{j,j} = \int_a^b \phi_j^2(x)dx$ in advance, and calculate
  $$\tilde{b}_j = \int_a^b f(x)\phi_j(x)dx.$$
  2. The sought coefficients are
  $$c_j = \tilde{b}_j / d_j, \quad j = 0, 1, \ldots, n.$$

- Next, restrict to polynomial best approximation and ask, how to construct a family of orthogonal polynomials?
- ASIDE: Two square-integrable functions $g \in L_2$ and $h \in L_2$ are orthogonal to each other if their inner product vanishes, i.e., they satisfy
$$\int_a^b g(x)h(x)dx = 0.$$

Best least squares using orthogonal basis functions

- The **construction algorithm** becomes simply:
  1. For $j = 0, 1, \ldots, n$, set $d_j = \tilde{B}_{j,j} = \int_a^b \phi_j^2(x)dx$ in advance, and calculate
  $$\tilde{b}_j = \int_a^b f(x)\phi_j(x)dx.$$
  2. The sought coefficients are
  $$c_j = \tilde{b}_j / d_j, \quad j = 0, 1, \ldots, n.$$

- Next, restrict to polynomial best approximation and ask, how to construct a family of orthogonal polynomials?
- ASIDE: Two square-integrable functions $g \in L_2$ and $h \in L_2$ are orthogonal to each other if their inner product vanishes, i.e., they satisfy
$$\int_a^b g(x)h(x)dx = 0.$$

Best least squares using orthogonal basis functions

- The **construction algorithm** becomes simply:
  1. For $j = 0, 1, \ldots, n$, set $d_j = \tilde{B}_{j,j} = \int_a^b \phi_j^2(x)dx$ in advance, and calculate
  $$\tilde{b}_j = \int_a^b f(x)\phi_j(x)dx.$$
  2. The sought coefficients are
  $$c_j = \tilde{b}_j / d_j, \quad j = 0, 1, \ldots, n.$$

- Next, restrict to polynomial best approximation and ask, how to construct a family of orthogonal polynomials?
- ASIDE: Two square-integrable functions $g \in L_2$ and $h \in L_2$ are orthogonal to each other if their inner product vanishes, i.e., they satisfy

$$\int_a^b g(x)h(x)dx = 0.$$

Legendre polynomials

Defined on the interval $[-1, 1]$ by

$$\begin{aligned}
\phi_0(x) &= 1, \\
\phi_1(x) &= x, \\
\phi_{j+1}(x) &= \frac{2j+1}{j+1} x \phi_j(x) - \frac{j}{j+1} \phi_{j-1}(x), \quad j \geq 1.
\end{aligned}$$

So,

$$\begin{aligned}
\phi_2(x) &= \frac{1}{2}(3x^2 - 1), \\
\phi_3(x) &= \frac{5}{6} x(3x^2 - 1) - \frac{2}{3} x = \frac{1}{2}(5x^3 - 3x).
\end{aligned}$$

Legendre polynomials

Defined on the interval $[-1, 1]$ by

$$\begin{aligned} \phi_0(x) &= 1, \\ \phi_1(x) &= x, \\ \phi_{j+1}(x) &= \frac{2j+1}{j+1} x\phi_j(x) - \frac{j}{j+1}\phi_{j-1}(x), \quad j \geq 1. \end{aligned}$$

So,

$$\begin{aligned} \phi_2(x) &= \frac{1}{2}(3x^2 - 1), \\ \phi_3(x) &= \frac{5}{6}x(3x^2 - 1) - \frac{2}{3}x = \frac{1}{2}(5x^3 - 3x). \end{aligned}$$

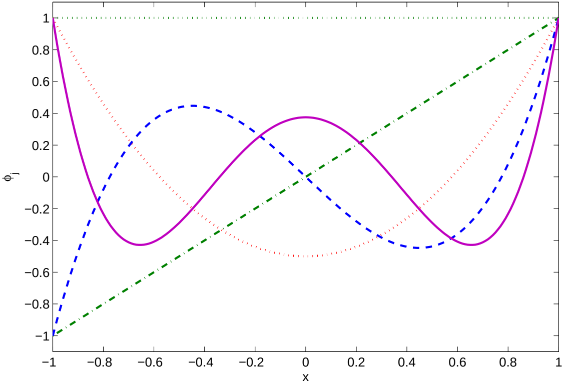
Legendre polynomials: properties

- *Orthogonality*:

$$\int_{-1}^{1} \phi_j(x)\phi_k(x)dx = \begin{cases} 0 & j \neq k \\ \frac{2}{2j+1} & j = k. \end{cases}$$

- *Calibration*: $|\phi_j(x)| \leq 1, \quad -1 \leq x \leq 1$, and $\phi_j(1) = 1$.
- *Oscillation*: $\phi_j(x)$ has degree j (not less). All its j zeros are simple and lie inside the interval $(-1, 1)$. Hence the polynomial oscillates j times in this interval.

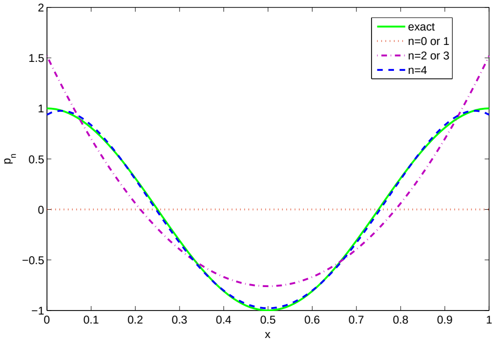
Legendre polynomials: the picture



You can determine which curve corresponds to which ϕ_j by the number of oscillations.

Example

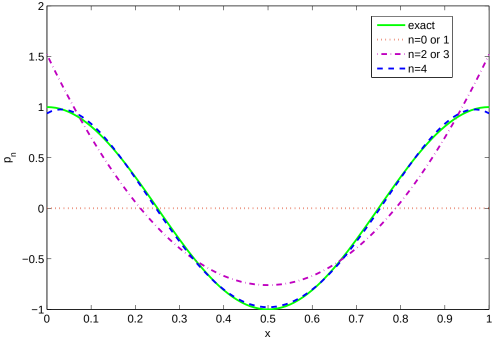
- Find polynomial best approximations for $f(x) = \cos(2\pi x)$ over $[0, 1]$ using n orthogonal polynomials.
- Using Legendre polynomials (note interval change from $[-1, 1]$ to $[0, 1]$), $\tilde{b}_j = \int_0^1 \cos(2\pi t)\phi_j(2t-1)dt$, and $c_j = (2j+1)\tilde{b}_j$.



Here f is in green. Obtain the same approximation for $n = 2l$ and $n = 2l + 1$ due to symmetry of f. Note improvement in error as l increases.

Example

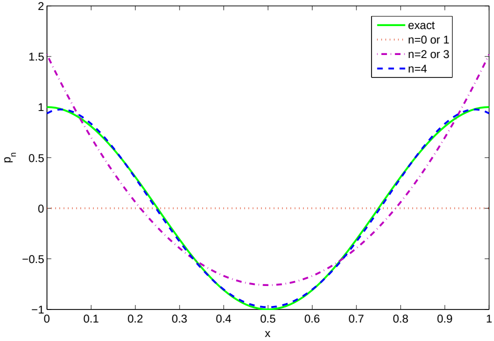
- Find polynomial best approximations for $f(x) = \cos(2\pi x)$ over $[0, 1]$ using n orthogonal polynomials.
- Using Legendre polynomials (note interval change from $[-1, 1]$ to $[0, 1]$), $\tilde{b}_j = \int_0^1 \cos(2\pi t)\phi_j(2t-1)dt$, and $c_j = (2j+1)\tilde{b}_j$.



Here f is in green. Obtain the same approximation for $n = 2l$ and $n = 2l + 1$ due to symmetry of f. Note improvement in error as l increases.

Example

- Find polynomial best approximations for $f(x) = \cos(2\pi x)$ over $[0, 1]$ using n orthogonal polynomials.
- Using Legendre polynomials (note interval change from $[-1, 1]$ to $[0, 1]$), $\tilde{b}_j = \int_0^1 \cos(2\pi t)\phi_j(2t-1)dt$, and $c_j = (2j+1)\tilde{b}_j$.



Here f is in green. Obtain the same approximation for $n = 2l$ and $n = 2l+1$ due to symmetry of f. Note improvement in error as l increases.

Trigonometric polynomials

- Not all interesting orthogonal basis functions are polynomials.
- Of particular interest are the periodic basis functions

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}}, \;\; \phi_{2l-1}(x) = \frac{1}{\sqrt{\pi}}\sin(lx), \; \phi_{2l}(x) = \frac{1}{\sqrt{\pi}}\cos(lx) \quad l = 1, 2, \ldots n/2.$$

- These are orthonormal on $[-\pi, \pi]$:

$$\int_{-\pi}^{\pi} \phi_i(x)\phi_j(x)dx = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}.$$

- Lead to the famous Fourier transform (Chapter 13).

Trigonometric polynomials

- Not all interesting orthogonal basis functions are polynomials.
- Of particular interest are the periodic basis functions

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}}, \;\; \phi_{2l-1}(x) = \frac{1}{\sqrt{\pi}}\sin(lx), \; \phi_{2l}(x) = \frac{1}{\sqrt{\pi}}\cos(lx) \quad l = 1, 2, \ldots n/2.$$

- These are orthonormal on $[-\pi, \pi]$:

$$\int_{-\pi}^{\pi} \phi_i(x)\phi_j(x)dx = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}.$$

- Lead to the famous Fourier transform (Chapter 13).

Outline

- Best least squares approximation
- Orthogonal basis functions
- Weighted least squares
- Chebyshev polynomials

Weighted least squares

- An important generalization of best least squares approximation: with a weight function $w(x)$ that is never negative, has bounded L_2 norm, and may equal 0 only at isolated points, seek coefficients c_j for $v(x) = \sum_{j=0}^{n} c_j \phi_j(x)$ so as to minimize

$$\int_a^b w(x)(v(x) - f(x))^2 dx.$$

- The orthogonality condition correspondingly generalizes to

$$\int_a^b w(x)\phi_j(x)\phi_k(x)dx = 0, \quad j \neq k.$$

- If this holds then the best approximation solution is given by $c_j = \tilde{b}_j / d_j$, where $d_j = \int_a^b w(x)\phi_j^2(x)dx > 0$ and

$$\tilde{b}_j = \int_a^b w(x)f(x)\phi_j(x)dx, \quad j = 0, 1, \ldots, n.$$

- Important families of orthogonal polynomials may be obtained for good choices of the weight function.

Weighted least squares

- An important generalization of best least squares approximation: with a weight function $w(x)$ that is never negative, has bounded L_2 norm, and may equal 0 only at isolated points, seek coefficients c_j for $v(x) = \sum_{j=0}^{n} c_j \phi_j(x)$ so as to minimize

$$\int_a^b w(x)(v(x) - f(x))^2 dx.$$

- The orthogonality condition correspondingly generalizes to

$$\int_a^b w(x)\phi_j(x)\phi_k(x)dx = 0, \quad j \neq k.$$

- If this holds then the best approximation solution is given by $c_j = \tilde{b}_j / d_j$, where $d_j = \int_a^b w(x)\phi_j^2(x)dx > 0$ and

$$\tilde{b}_j \;=\; \int_a^b w(x)f(x)\phi_j(x)dx, \quad j = 0, 1, \ldots, n.$$

- Important families of orthogonal polynomials may be obtained for good choices of the weight function.

Gram-Schmidt process for constructing orthogonal polynomials

- Given an interval $[a, b]$ and a weight function $w(x)$, this algorithm produces an orthogonal polynomial family via a short three-term recurrence!
- Set

$$\begin{aligned}\phi_0(x) &= 1,\\ \phi_1(x) &= x - \beta_1,\\ \phi_j(x) &= (x - \beta_j)\phi_{j-1}(x) - \gamma_j \phi_{j-2}(x), \quad j \geq 2,\end{aligned}$$

where

$$\begin{aligned}\beta_j &= \frac{\int_a^b x w(x)[\phi_{j-1}(x)]^2 dx}{\int_a^b w(x)[\phi_{j-1}(x)]^2 dx}, \quad j \geq 1,\\ \gamma_j &= \frac{\int_a^b x w(x)\phi_{j-1}(x)\phi_{j-2}(x) dx}{\int_a^b w(x)[\phi_{j-2}(x)]^2 dx} \quad j \geq 2.\end{aligned}$$

- Intimately connected also to the derivation of Krylov subspace iterative methods for symmetric linear systems (CG, MINRES); see Section 7.5.

Outline

- Best least squares approximation
- Orthogonal basis functions
- Weighted least squares
- Chebyshev polynomials

Chebyshev polynomials

- This is a family of orthogonal polynomials on $[-1, 1]$ with respect to the weight function

$$w(x) = \frac{1}{\sqrt{1-x^2}}.$$

- Define

$$\phi_j(x) = T_j(x) = \cos(j \;\arccos x) = \cos(j\theta),$$

where $x = \cos(\theta)$.

- So, these polynomials are naturally defined in terms of an angle, θ. Clearly, the larger j the more oscillatory T_j.
- Easy to verify orthogonality:

$$\begin{aligned}\int_{-1}^{1} w(x)T_j(x)T_k(x)dx &= \int_{-1}^{1} \frac{T_j(x)T_k(x)}{\sqrt{1-x^2}}dx \\ &= \int_0^{\pi} \cos(j\theta)\cos(k\theta)d\theta = \begin{cases} 0 & j \neq k \\ \frac{\pi}{2} & j = k > 0 \end{cases}.\end{aligned}$$

Chebyshev polynomials: derivation and properties

- These polynomials obey the simple 3-term recursion

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{j+1}(x) &= 2xT_j(x) - T_{j-1}(x), \quad j \geq 1, \end{aligned}$$

  and obviously satisfy

$$|T_j(x)| \leq 1, \quad \forall x.$$

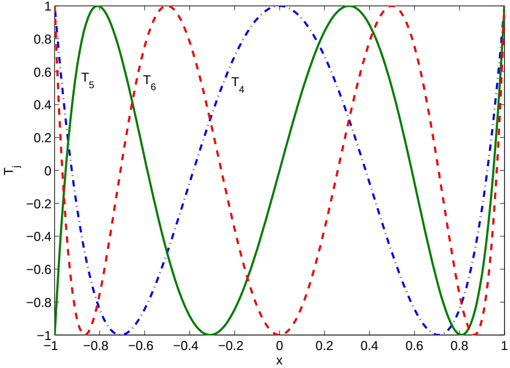
- The roots (zeros) of T_n are the Chebyshev points

$$x_k = \cos\left(\frac{2k-1}{2n}\pi\right), \quad k = 1, \ldots, n.$$

  See Section 10.6 for their magic properties.
- The also-important interleaving extremal points (where $T_n(x) = \pm 1$) are

$$\xi_k = \cos\left(\frac{k}{n}\pi\right), \quad k = 0, 1, \ldots, n.$$

Chebyshev polynomials: the picture



You can determine which curve corresponds to which $\phi_j = T_j$ by the number of oscillations. Note the perfect calibration.