FALL 200

P.L. CHEBYSHEV AND CHEBYSHEV POLYNOMIALS

J. ANTHONY COOKSON

Montana State University

Department of Mathematical Sciences

Statistics Writing Project

ABSTRACT. After a short survey of the mathematical contributions of P.L. Chebyshev, this paper presents a mathematical formulation of Chebyshev polynomials, an especially useful kind of orthogonal polynomials. To demonstrate one use of Chebyshev polynomials, the paper describes the problem of finding the polynomial of order n that minimizes the squared residual distance of m data points to the fitted function. If the standard basis for $\mathbb{R}^n$ is used, this estimation problem is plagued by multicollinearity, which results in numerical instability of the estimates. Chebyshev polynomials provide a numerically stable way of estimating the least-squares solution. This paper presents an algorithm for the use of Chebyshev polynomials to obtain a least-squares fit and works through an example to demonstrate the method outlined in this paper.

1. Introduction

"Chebyshev polynomials are everywhere dense in numerical analysis." This quote, which Sarra (2006) attributes to "a number of distinguished mathematicians," illustrates the richness and detail that characterize the study of Chebyshev polynomials. Applications of Chebyshev polynomials permeate numerical analysis and mathematics in general, making the potential scope of a paper on the applications of Chebyshev polynomials enormous. For this reason, this paper does not attempt to provide a complete survey of numerical methods that rely on Chebyshev polynomials. Rather, this paper aims to discuss adequately one such application, polynomial approximation to data using a least squares criterion, while pointing the interested reader toward other applications of Chebyshev polynomials.

The discussion of Chebyshev polynomials is the heart of this paper, but this paper cannot do justice to the study of Chebyshev polynomials without recounting important details of Pafnuty Chebyshev's life and academic career. This paper, therefore, begins by providing context, describing Chebyshev's life and notable accomplishments in Section 2. Section 3 delves into the mathematics of Chebyshev polynomials, presenting a mathematical formulation, including their important properties as sets of orthogonal polynomials. Section 4 describes the least squares problem and presents an algorithm to find the least squares approximation to data using Chebyshev polynomials. Section 5 surveys some other applications of Chebyshev polynomials, points to current research agendas, describing an important weakness in Chebyshev methods – inability to effectively deal with discontinuous functions. Lastly, Section 6 concludes with the overall message of this paper.

2. Historical Information[1]

Pafnuty Lvovich Chebyshev was born on May 16, 1821 in Okatovo, Russia to an upper middle class military family. Shortly before Chebyshev's birth, Russia had just defeated Napoleon and established itself as a world power. This new position on the world stage sparked debate among Russians about Russia's relationship to Europe. Some nationals argued that Russia should isolate itself because they viewed other nations as inferior. Others saw the benefit of cultural and economic exchange and, therefore, pushed for the westernization of Russia. This latter group was

[1]This section is an adaptation from O'Connor and Robertson (2006). See [6].

mostly comprised of military families who were acquainted with European cultures and traditions. Because Chebyshev came from a military family, he was part of this second movement. Therefore, Chebyshev was brought up to appreciate European culture and was tutored in French by his cousin, who also taught him arithmetic. Chebyshev's fluency in French would become one of his greatest assets in establishing himself as a mathematician on the world stage.

2.1. **Education and Early Work.** In 1832, Chebyshev's family moved to Moscow. In Moscow, Chebyshev was tutored by P.N. Pogorelski, a man some considered to be the best elementary mathematics tutor in Moscow. This early influence gave Chebyshev a solid mathematical foundation and inspired him to take up mathematics as a career. Chebyshev was, therefore, well-prepared for university study at Moscow University. There, under Nikolai D. Brashman, he was exposed to mechanical engineering, hydraulics, the theory of integration of algebraic functions, as well as the calculus of probability. These early studies served as a springboard for Chebyshev's research interests. He would later make large contributions to each of these fields.

Chebyshev aspired to international fame as a mathematician. Late in his career, he even objected to being called a "splendid Russian mathematician," insisting that the scope of his influence was much more widespread than Russia. Indeed, Chebyshev's significance spread well beyond Russia; Chebyshev's contributions retain their significance to this day. His personal drive to become a world-wide mathematician colored his early work and the way he presented his findings. His first publication was published in Liouville's journal in France in 1842. As France was the world center of mathematics at the time, this first publication in a French journal was a tremendous accomplishment and a sign that Chebyshev was emerging as a great mathematician.

2.2. **Breadth and Scope of Contributions to Mathematics.** Chebyshev made numerous and wide-ranging contributions to mathematics. In probability, he is best known for the Chebyshev inequality, which is used for a convenient proof for the weak law of large numbers. He also pioneered work on random variables and expected values, providing much of the foundation for the application of probability to statistical data.

In number theory, he proved several important results regarding prime numbers including Bertrand's conjecture that for $n \in \mathbb{N}$ where $n > 3$, $\exists$ at least one prime number, p, so that $n < p < 2n$. In addition, Chebyshev was instrumental in producing a complete edition of Euler's 99 number theory papers, published in 1849. Chebyshev also made contributions to the theory of integrals (generalizing the beta function), the construction of maps, the calculation of geometric volumes, and the construction of calculating machines. In addition, Chebyshev was also an inventor. In 1893, a year before Chebyshev's death, the World's Exposition in Chicago exhibited seven of his mechanical inventions, one of which was a special bicycle for women.

2.3. **Contribution to Numerical Analysis: Chebyshev Polynomials.** Chebyshev's work on approximation theory appears to have been largely inspired by a trip in 1852 through Western Europe where he examined various steam engines and their mechanics in practice. Tikhomirov later applied Chebyshev's work on approximation theory to the theory of mechanisms, noting that Chebyshev's work on the subject inspired his own. In 1854, Chebyshev published Théorie des mécanismes connus sous le nom de parallélogrammes, which was the first work to feature Chebyshev polynomials. He would later expound greatly on those initial ideas, generalizing the idea of orthogonal polynomials.

Chebyshev was one of the first mathematicians to recognize the power of orthogonal polynomials. Others, namely Legendre, had discovered applications of orthogonal polynomials before Chebyshev's time, but Chebyshev greatly expanded what was known, as well as the range of applications regarding orthogonal polynomials. Specifically, Chebyshev developed a general theory of orthogonal polynomials, for which he is well known. His work in the field of approximation theory was so vast that many of his contributions went unheralded until well after his death. Of note in this regard are Hahn Polynomials and the Christoffel-Darboux formula [8].

3. Mathematical Formulation

The remainder of this paper is concerned with Chebyshev polynomials and their application to the least squares problem. Chebyshev polynomials are a special case of orthogonal polynomials. Therefore, it is important to consider properties of orthogonal polynomials first to provide context for the discussion of Chebyshev polynomials.

3.1. **Orthogonal Polynomials.** Orthogonal polynomials have widespread application in mathematical and physical sciences because they provide a natural means of approximating functions and, therefore, ease in solving complicated differential equations. The general application of orthogonal polynomials to solve differential equations is, however, beyond the scope of this paper. More to the point, this section presents the more useful and interesting results relating orthogonal polynomials, leaving further investigation to the reader. Before describing some of the more interesting results on orthogonal polynomials, three definitions are necessary.

Definition 3.1. Two polynomials, $p(x)$ and $q(x)$ are **orthogonal** on $[a, b]$ if

$$\int_a^b w(x)p(x)q(x)dx = 0 \qquad (1)$$

where $w(x)$ is a weighting function that satisfies $w(x) > 0, \forall x \in (a, b)$.

Definition 3.2. A set of polynomials $P = \{p_0, p_1, ..., p_n\}$, is an **orthogonal set** on $[a, b]$ if $p_i(x)$ and $p_j(x)$ are orthogonal for all $i \neq j$.

The following definition implements this notion of orthogonal sets of polynomials

Definition 3.3. The sequence of polynomials $\{p_n\}$ is called an **orthogonal sequence** of polynomials on $[a, b]$ if every subsequence of the form $P = \{p_0, p_1, ..., p_n\}$ forms an orthogonal set on $[a, b]$.

Here $[a, b]$ is called the **interval of orthogonality**, which is the interval on which (1) holds. Two notable properties of orthogonal polynomials are the *existence of real roots property* and the *interlacing of roots property*. There is also a general *recurrence formula*, which is useful for finding orthogonal polynomials in terms of the previous two terms in the orthogonal sequence. These properties are stated and proven in the following four theorems. Proofs of these theorems were adapted from the references (See [5] for alternate proofs).

Theorem 3.4. *If p_n be the n^{th} order term in an orthogonal sequence of polynomials with interval of orthogonality $[a, b]$, then p_n is orthogonal to any polynomial of lower degree on the same interval of orthogonality.*

Proof. Let $S(x)$ be a polynomial degree less than or equal to $n-1$.

Intuitively, there exist constants α_i such that $S(x) = \sum_{i=0}^{n-1} \alpha_i p_i$. Using this fact in the first step, form the inner product over the integral of orthogonality:

$$\begin{aligned}
\int_a^b S(x)p_n(x)w(x)dx &= \int_a^b \left(\sum_{i=0}^{n-1} \alpha_i\right) p_i p_n(x)w(x)dx \\
&= \sum_{i=0}^{n-1} \alpha_i \int_a^b p_i p_n(x)w(x)dx \\
&= 0
\end{aligned}$$

The second line is valid because the summation is a finite summation, and can, therefore, be pulled out of the integral. The last line uses the orthogonality of p_n to every lower-order term in the orthogonal sequence. Thus, we have that $S(x)$ and $p_n(x)$ are orthogonal if $S(x)$ is a polynomial of a lower degree than $p_n(x)$. □

Theorem 3.5. *(existence of real roots): Each polynomial in an orthogonal sequence $\{p_k\}_{k=0}^{\infty}$ has k distinct, real roots, all of which lie strictly within the interval of orthogonality $[a, b]$.*

Proof. Let m be the number of times that the polynomial $p_n(x)$ changes sign inside the interval of orthogonality. Let $R = \{x_1, x_2, ..., x_m\}$ be the set of the points where $p_n(x)$ changes sign. By definition of root, R is the set of roots of the polynomial $p_n(x)$. By the fundamental theorem of algebra, we know that $m \leq n$. The only way that $m = n$ is if the all of the roots of $p_n(x)$ are distinct, real, and lie inside the interval of orthogonality. Thus, the proof is done if we can show that $m = n$.

Let $S(x) = \prod_{i=1}^{m}(x - x_i)$, where x_i, $i = 1, ..., m$ are the roots of $p_n(x)$. Notice that $S(x)$ is a m^{th} degree polynomial that changes sign whenever $p_n(x)$ changes sign within the interval of orthogonality. This means that $\forall x \in R^c \cap [a, b]$, $S(x)$ and $p_n(x)$ either have the same sign as one another (Case I) or have opposite sign from one another (Case II). In Case I, $S(x)p_n(x) > 0 \forall x \in R^c \cap [a, b]$. In Case II, $S(x)p_n(x) < 0 \ \forall x \in R^c \cap [a, b]$.

Now, consider the inner product between $S(x)$ and $p_n(x)$, where $w(x) > 0$ is the weighting function:

$$\int_a^b S(x)p_n(x)w(x)dx \neq 0$$

Under Case I, this integral is positive. Under Case II, this integral is negative. Regardless, it is not zero. Thus, we know that $S(x)$ and $p_n(x)$ are not orthogonal. By the previous theorem, this implies that $S(x)$ is a polynomial of degree n or larger. Clearly, it cannot be larger. Thus, $m = n$.

□

Theorem 3.6. *Theorem 3.6. (recurrence formula for orthogonal sequences): If $\{p_k\}_{k=0}^{\infty}$ is an orthogonal sequence with interval of orthogonality $[d_1, d_2]$, then $p_{n+1} = (a_n x + b_n)p_n - c_n p_{n-1}$ where the coefficients a, b, and c depend on n.*

Proof. Choose a so that the x^{n+1} terms of $axp_n(x)$ and p_{n+1} match. Then, we know that $axp_n(x) - p_{n+1} = S_n(x)$, a polynomial with degree n or less. We can also choose b, so that the x^n terms of $(ax+b)p_n(x)$ and p_{n+1} match. This implies $(a+bx)p_n(x) - p_{n+1} = S_{n-1}(x)$, a polynomial with degree $n-1$ or less. Expressing $S_{n-1}(x)$ as a linear combination of the polynomials in the orthogonal sequence, the previous expression reduces to:

$$(ax+b)p_n(x) - p_{n+1}(x) = \sum_{i=1}^{n-1} \lambda_i p_i(x) \tag{2}$$

Now, multiply both sides of this expression by $p_j(x)w(x)$, where $p_j(x)$ is the j^{th} $(j \leq n-2)$ polynomial in the orthogonal sequence and $w(x)$ is the weighting function for the orthogonal sequence.

$$\begin{aligned}
(ax+b)p_n(x)p_j(x)w(x) - p_{n+1}(x)p_j(x)w(x) &= \left(\sum_{i=1}^{n-1} \lambda_i p_i(x)\right) p_j(x)w(x) \\
\begin{array}{c} bp_n(x)p_j(x)w(x)+ \\ axp_n(x)p_j(x)w(x) - p_{n+1}(x)p_j(x)w(x) \end{array} &= \left(\sum_{i\neq j}^{n-1} \lambda_i p_i(x)p_j(x)w(x)\right) + \lambda_j p_j(x)p_j(x)w(x)
\end{aligned}$$

If we take the integral over the interval of orthogonality of both sides of this expression , the equality remains:

$$\begin{array}{rcl} b\int_{d_1}^{d_2} p_n(x)p_j(x)w(x)dx + & & \left(\sum_{i\neq j}^{n-1} \lambda_i \int_{d_1}^{d_2} p_i(x)p_j(x)w(x)dx\right) + \\ a\int_{d_1}^{d_2} xp_n(x)p_j(x)w(x)dx - \int_{d_1}^{d_2} p_{n+1}(x)p_j(x)w(x) & = & \lambda_j \int_{d_1}^{d_2} p_j(x)p_j(x)w(x)dx \\ 0 + a\int_{d_1}^{d_2} p_n(x)\left(xp_j(x)\right)w(x)dx - 0 & = & (0) + \lambda_j \int_{d_1}^{d_2} p_j(x)p_j(x)w(x)dx \end{array}$$

where going from the first expression to the second expression is accomplished by recognizing that $p_k(x)$ and $p_l(x)$ are orthogonal for $k \neq l$. Notice the rearrangement of the terms inside the integral on the left hand side of the expression. We know that $xp_j(x)$ is a polynomial of degree no greater than $n-1$ because $p_j(x)$ is a polynomial of degree no greater than $n-2$. Thus, by the first theorem in this section, $p_n(x)$ and $xp_j(x)$ are orthogonal and the integral on the left hand side is zero. Now, notice that for $j \leq n-2$, we have:

$$\lambda_j \int_{d_1}^{d_2} p_j(x)p_j(x)w(x)dx = 0$$

The integral in this expression is clearly positive, so λ_j must equal zero for $j \leq n-2$. This implies that the only λ_j that can be nonzero is λ_{n-1}. Thus, equation (2) reduces to:

$$(ax+b)p_n(x) - p_{n+1}(x) = \lambda_{n-1}p_{n-1}(x)$$

Set $c = \lambda_{n-1}$ and we have:

$$p_{n+1}(x) = cp_{n-1}(x) - (ax+b)p(x),$$

which is the desired result. □

Remark 3.7. Given the coefficients the first couple of terms, there is a closed form for each of a, b, and c. For orthogonal polynomials with an interval of orthogonality of $[d_1, d_2]$, let k_j^i be the coefficient on the i^{th} order term for the j^{th} order orthogonal polynomial in the sequence. Then, the following are closed form representations:

$$
\begin{aligned}
a &= \frac{k_{n+1}^{n+1}}{k_n^n} \\
b &= a\left(\frac{k_{n+1}^{n}}{k_{n+1}^{n+1}} - \frac{k_n^{n-1}}{k_n^n}\right) \\
c &= a\left(\frac{k_{n-1}^{n-1}\int_{d_1}^{d_2}[p_n(x)]^2\,w(x)dx}{k_n^n\int_{d_1}^{d_2}[p_{n-1}(x)]^2\,w(x)dx}\right)
\end{aligned}
$$

Lemma 3.8. *Suppose that $p_n(x)$ and $p_{n+1}(x)$ are terms in a sequence of polynomials $\{p_k\}_{k=0}^{\infty}$ that are orthogonal on $[a, b]$. Suppose also that the leading term in each polynomial in the sequence is positive. Then:*

$$p'_{n+1}(x)p_n(x) \;>\; p_{n+1}(x)p'_n(x)$$

for any $x \in [a, b]$, where p'_k means $\frac{dp_k(x)}{dx}$.

Proof. (by induction). First, establish the base case: Let $n = 0$. Clearly, $p'_1(x) > 0$ and $p_0(x) > 0$ because the leading terms in each of the polynomials is greater than zero. We also know that $p'_0(x) = 0$ because $p_0(x)$ is constant with respect to x. Thus, it is true that:

$$p'_{n+1}(x)p_n(x) \;>\; p_{n+1}(x)p'_n(x) = 0$$

Now, establish the general case (assume the statement holds for n, show that it also holds for $n+1$): Use the recurrence formula (the previous theorem) to get an expression for $p_{n+1}(x)$:

$$
\begin{aligned}
p_{n+1}(x) &= (ax+b)p_n(x) - cp_{n-1}(x) \\
\Rightarrow p'_{n+1}(x) &= ap_n(x) + (ax+b)p'_n(x) - cp'_{n-1}(x)
\end{aligned}
$$

Now, form the expression:

$$
\begin{aligned}
p'_{n+1}(x)p_n(x) - p_{n+1}(x)p'_n(x) &= \left[ap_n(x) + (ax+b)p'_n(x) - cp'_{n-1}(x)\right] p_n(x) - \\
&\qquad \left[(ax+b)p_n(x) - cp_{n-1}(x)\right] p'_n(x) \\
&= ap_n^2(x) + (ax+b)p'_n(x)p_n(x) - (ax+b)p'_n(x)p_n(x) \\
&\qquad -cp'_{n-1}(x)p_n(x) + cp_{n-1}(x)p'_n(x)
\end{aligned}
$$

Then, some cancelation yields

$$
\begin{aligned}
p'_{n+1}(x)p_n(x) - p_{n+1}(x)p'_n(x) &= ap_n^2(x) + c\left(p'_n(x)p_{n-1}(x) - p_n(x)p'_{n-1}(x)\right) \\
&\geq c\left(p'_n(x)p_{n-1}(x) - p_n(x)p'_{n-1}(x)\right)
\end{aligned}
$$

Now, by the induction hypothesis, we know $p'_n(x)p_{n-1}(x) - p_n(x)p'_{n-1}(x) > 0$. From Remark 3.7, we know that $c = \frac{k_{n+1}^{n+1}}{k_n^n}\left(\frac{k_{n-1}^{n-1}\int_{d_1}^{d_2}[p_n(x)]^2 w(x)dx}{k_n^n \int_{d_1}^{d_2}[p_{n-1}(x)]^2 w(x)dx}\right)$. This expression is greater than zero because it is a function of the leading terms of the polynomials (which are greater than zero by hypothesis) and the integral of a squared polynomial times a positive weighting function. This implies that $p'_{n+1}(x)p_n(x) - p_{n+1}(x)p'_n(x) > 0$, which is the desired result. □

Theorem 3.9. *(interlacing of roots): Consider a sequence of polynomials* $\{p_k\}_{k=0}^{\infty}$, *orthogonal on the interval* $[a,b]$. *Let* $\{x_1 < x_2 < ... < x_n\}$ *be the roots of* $p_n(x)$ *with* $x_0 = a$ *and* $x_{n+1} = b$. *For notational convenience, let* $\{y_1 < y_2 < ... < y_n < y_{n+1}\}$ *be the roots of* p_{n+1}. *Then, each interval of the form* (x_j, x_{j+1}) *for* $j = 0, 1, ..., n$, *will contain exactly one of the roots of* $p_{n+1}(x)$.

Proof. Suppose that y_j is a root of $p_{n+1}(x)$. Lemma 3.7 implies that at this value:

$$p'_{n+1}(y_j)p_n(y_j) \;>\; p_{n+1}(y_j)p'_n(y_j) = 0$$

This inequality implies that $p'_{n+1}(x)$ and $p_n(x)$ have the same sign at any root of $p_{n+1}(x)$. It also implies that any root for $p_{n+1}(x)$ cannot be a root of $p_n(x)$ (otherwise the inequality would not be strict). By Rolle's Theorem, there must be a point between y_j and y_{j+1} where $p'_{n+1}(x)$ changes sign. Thus, $p_n(x)$ changes sign in the open interval (y_j, y_{j+1}). Call this point x_j.

So far, we've established that there exists a root of $p_n(x)$ such that $x_j \in (y_j, y_{j+1})$ for every interval where the endpoints roots of p_{n+1}. To establish the result that there exists exactly one root of $p_{n+1}(x), y_j \in (x_j, x_{j+1})$, we need to establish that $p_n(x)$ changes sign only once between roots of $p_{n+1}(x)$. To establish this fact, suppose that there are two roots of $p_n(x)$ in the interval (y_j, y_{j+1}). Then, because there is at least one root of $p_n(x)$ in the intervals $(y_1, y_2), (y_2, y_3), ..., (y_n, y_{n+1})$. If any of these n intervals has more than one root for $p_n(x)$, then $p_n(x)$ would have more roots than its highest order term, a contradiction.

Thus, there is exactly one root of p_n in each of the intervals $(y_1, y_2), (y_2, y_3); ..., (y_n, y_{n+1})$. By defining $x_0 = a$, and $x_{n+1} = b$, we can say that $y_1 \in (x_0, x_1)$ and that $y_{n+1} \in (x_n, x_{n+1})$. We can, therefore, turn the statement around and say that there exists exactly one root of $p_{n+1}(x)$ in each interval of the form (x_j, x_{j+1}) where $j = 1, 2, ..., n$. □

3.2. Overview of Chebyshev Polynomials[2].

This section begins with a definition:

Definition 3.10. A set of Chebyshev Polynomials[3] $\{T_k(x)\}_{k=0}^{k=n}$ is a set of polynomials satisfying (1) with respect to the weighting function

(3) $$w(x) = \frac{1}{\sqrt{1+x^2}}$$

[2]This section draws largely from material on the Math-World Web Resource. See [8].

[3]This paper uses "Chebyshev polynomials" to refer to Chebyshev polynomials of the first kind. There are also Chebyshev polynomials of the second kind, which are not discussed in this paper. For more information, see [8].

The interval of orthogonality for Chebyshev Polynomials is $[-1,1]$. More usefully, Chebyshev polynomials can be defined through the the following identity involving cosines:

$$T_n(cos(\theta)) = cos(n\theta) \tag{4}$$

If we let x=cos(θ), expression (4) becomes:

$$T_n(x) = cos(n(cos^{-1}(x))) \tag{5}$$

Sequences of Chebyshev polynomials can also be generated recursively with the following two recurrence relations.

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{6}$$

$$T_{n+1}(x) = xT_n(x) - \sqrt{(1-x^2)(1-(T_n(x))^2)} \tag{7}$$

where $T_0(x) = 1$ and $T_1(x) = x$. The expression in (6) is simpler than the one in (7), but requires more knowledge in the sense that one must know the previous two polynomials in order to use (6), whereas (7) requires only knowledge of the previous polynomial.

Aside from $T_0 = 1$, the first five Chebyshev polynomials are as follows:

$$\begin{aligned}
T_1(x) &= x \\
T_2(x) &= 2x^2 - 1 \\
T_3(x) &= 4x^3 - 3x \\
T_4(x) &= 8x^4 - 8x^2 + 1 \\
T_5(x) &= 16x^5 - 20x^3 + 5x
\end{aligned}$$

FIGURE 1. The First Five Chebyshev Polynomials

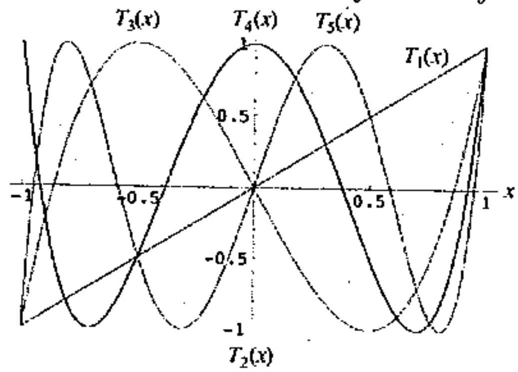

Figure 1 displays the first five Chebyshev polynomials, plotted on their interval of orthogonality $[-1, 1]$.[4]

Being a set of orthogonal polynomials, the roots of Chebyshev polynomials are distinct, real, are all in the interval $[-1, 1]$, and behave according to the interlacing of roots property. More specifically, there is a closed form expression for finding the roots of a Chebyshev polynomial, $T_n(x)$, as given by the following formula for $0 \leq k \leq n$. These roots are known as the Chebyshev-Gauss (CG) points [5].

$$x = cos(\frac{\pi(k - \frac{1}{2})}{n}) \tag{8}$$

The extrema for any given Chebyshev polynomial also have a simple representation for $0 < k < n$. The set of extrema taken together with the endpoints are called the Chebyshev-Gauss-Lobatto (CGL) points and are useful in numerical integration [5].

$$x = cos(\frac{k\pi}{n}) \tag{9}$$

In addition to those mentioned in this document, there are several other recursion formulas that can be used to generate Chebyshev polynomials. Chebyshev polynomials also have important relationships with multiple-angle formulas and have representations as complex integrals. There are

[4]This figure was taken from [6].

some interesting determinant relationships and a vast set of applications for Chebyshev polynomials. In addition, there are important relationships between Chebyshev polynomials and Bessel functions. More information along these lines can be found in the references (See [5, 6]).

4. Least Squares and Chebyshev Polynomials

Chebyshev polynomials have broad application in applied mathematics and physical sciences. One especially important application of Chebyshev polynomials within the field of numerical analysis is the application of Chebyshev polynomials to the least squares problem. To provide appropriate context for this application, this section defines and describes the least squares problem before applying the Chebyshev polynomial algorithm for approximating the least squares solution to the normal equations.

4.1. The Least Squares Problem. Much of statistical theory seeks to find a "best approximation" to real world data. In the world of statistics, the most often employed criterion to gauge what constitutes the "best" approximation by the criterion of least squares. Consider the linear case first. That is, suppose $\{y_k\}_{k=0}^{m}$ is a set of data points that we wish to approximate with a linear equation, $g(x) = a + bx$. For linear regression, the least squares approximation is the one that minimizes (with respect to a and b).

$$\psi(a,b) = \sum_{k=0}^{m}(ax_k + b - y_k)^2 \tag{10}$$

Differentiating (10) with respect to a and b and setting the result equal to zero yields what are referred to as the *normal equations* for a simple linear function.

$$\begin{aligned}(\sum_{k=0}^{m} x_k^2)a + (\sum_{k=0}^{m} x_k)b &= \sum_{k=0}^{m} y_k x_k \\ (\sum_{k=0}^{m} x_k)a + (m+1)b &= \sum_{k=0}^{m} y_k\end{aligned}$$

The solution to the normal equations yields a and b that minimize sum of the squared deviations from the response and the fitted function, which in this case is a straight line.

4.2. Least Squares for Nonlinear Approximations. The method of least squares is not restricted to linear regression. In fact, many of the most interesting applications of the least squares criterion arise in nonlinear settings. Cheney and Kincaid (2003) illustrate how to obtain normal equations when searching for the best functional approximation of any particular form (e.g., exponential functions, log functions, etc.). In fact, they offer examples of obtaining normal equations when the functional form is not a polynomial, but a transcendental function (See [1]).

For numerical purposes, however, it is common to construct the functional approximation to the data using a set of polynomial basis functions for $\mathbb{R}^n$, where n is the highest order polynomial employed. In this case, the resulting fitted function is a polynomial of degree n. Using a set of functions, $\{g_i(x)\}_{k=0}^{n}$, which form a basis for $\mathbb{R}^n$, the least squares problem becomes min $\psi(c_1, c_2, ..., c_n)$ where:

$$\psi(c_1, c_2, ..., c_n) \quad = \quad \sum_{k=0}^{m}[\sum_{j=0}^{n} c_j g_j(x_k) - y_k]^2$$

The normal equations for this situation are obtained by differentiating with respect to c_i for $0 \leq i \leq n$ and equating the resulting linear combinations to zero. The i^{th} normal equation, therefore, becomes:

$$\sum_{j=0}^{n}[\sum_{k=0}^{m} g_i(x_k) g_j(x_k)]c_j \quad = \quad \sum_{k=0}^{m} y_k g_i(x_k)$$

One of the more natural choices of a basis for $\mathbb{R}^n$ is the set $\{1, x, x^2, ..., x^n\}$, which clearly spans $\mathbb{R}^n$. This set, however, is often a poor basis for numerical purposes because these functions are all very similar and pick up on similar behavior in the data. Employing orthonormal bases, like sets of Chebyshev polynomials, can correct this redundancy by ensuring that each of the polynomials in the set is different enough from the others. In this way, Chebyshev polynomials provide an ideal

Algorithm 1 Using Chebyshev polynomials to obtain the best approximation to a set of $m+1$ data points using a polynomial of order n or less.

(1) Let $a = min\{x_k\}$ and $b = max\{x_k\}$. Then $[a, b]$ is the smallest interval containing all the data points.
(2) Transform the data to the interval $[-1, 1]$ by using the change of variables $x_{new} = \frac{2x-a-b}{b-a}$. This transformation is important because Chebyshev polynomials are orthogonal polynomials with an interval of orthogonality of $[-1, 1]$. The useful properties of Chebyshev polynomials, therefore, provide a good approximation to the data on the interval $[-1, 1]$.
(3) Use Chebyshev polynomials to generate the $(n+1) \times (n+1)$ normal equations. These equations are given by $\sum_{j=0}^{n}[\sum_{k=0}^{m} T_i(x_k)T_j(x_k)]c_j = \sum_{k=0}^{m} y_k g_i(x_k)$ for $0 \leq i \leq n$.
(4) Solve for $c_0, c_1, ..., c_n$ using some equation solving routine.
(5) Use $c_0, c_1, ..., c_n$ to form $g(x) = \sum_{j=0}^{n} c_j T_j(x_{new})$, defined on the interval $[-1, 1]$.
(6) Transform the fitted function back to the interval $[a, b]$. The fitted polynomial is $g(\frac{2x-a-b}{b-a})$.

tool with which to address the problem of finding the best polynomial approximation to the real world data.

4.3. **Chebyshev Polynomials in a Least Squares Setting.** Cheney and Kincaid (2003) describe an algorithm for using linear combinations of polynomials to produce a polynomial of best fit to real world data. I repoduce this algorithm at the top of this page (Algorithm 1). This subsection provides some some additional details regarding why particular steps in the algorithm are needed.

The goal of the algorithm is to obtain the best approximation to a set of $m+1$ data points, $\{x_k\}_{k=0}^{m}$, using a polynomials order n or less, where $m > n$. Generally, it is not useful to fit high-order polynomials (high n) to data. Rather, it is desirable to choose the smallest n that does a reasonable job approximating the $m+1$ data points.

The application of Algorithm 1 is not limited to using Chebyshev polynomials. The algorithm allows for other orthogonal sets of polynomials to be used in place of Chebyshev polynomials. The algorithm does, however, produce better results with Chebyshev polynomials because they are a diffuse, orthogonal set of functions.

5. Scope of Application

Beyond their application to least squares, Chebyshev polynomials have broad application in mathematics. This section gives an example for one such application and discusses an important limitation of Chebyshev approximation techniques – Gibbs Phenomena.

5.1. Motivation for the Application. One particularly nice consequence of Algorithm 1 is that it yields a set of normal equations whose solution has better numerical stability than the standard approach. Numerical stability is a concern with the ordinary basis for $\mathbb{R}^n$ because the functions in the set $\{1, x, x^2, ..., x^n\}$ are highly collinear with one another. Where using the ordinary polynomial basis leads to problems with multicollinearity, the algorithm in the previous section harnesses the ortogonality of Chebyshev polynomials to circumvent these problems. The result is better numerical stability in the estimates.

The following definition helps to make this discussion more precise:

Definition 5.1. The **condition number** of a matrix **A** is defined to be $\kappa(\mathbf{A}) = \frac{\sigma_{max}(\mathbf{A})}{\sigma_{min}(\mathbf{A})}$ where $\sigma_{max}(\mathbf{A})$ and $\sigma_{min}(\mathbf{A})$ are minimal and maximal singular values of **A**,[5] respectively.

Matrices with small condition numbers have better numerical stability than those matrices with large condition numbers. A problem with a low condition number is said to be **well-conditioned**, whereas a problem with a high condition number is said to be **ill-conditioned**.

5.2. A Least Squares Polynomial Example. The following table of data (where x_i are values of the explanatory variable and y_i are values of the response variable) provide an illustrative example on the use of Chebyshev polynomials. In this example, we seek to fit a polynomial of order five to the nine data points.

i	1	2	3	4	5	6	7	8	9
x_i	1.0	1.1	1.3	1.5	1.9	2.1	2.4	2.8	3.5
y_i	1.84	1.96	2.21	2.45	2.94	3.18	4.9	5.2	5.7

The first two steps of Algorithm 1 tell us to transform the x_i to the interval of orthogonality for Chebyshev polynomials, $[-1, 1]$. Under this transformation, the data change to the following table.

[5]Every matrix has a singular value decomposition. For a proof of this fact, see [1].

i	1	2	3	4	5	6	7	8	9
x_i	-1.0	-0.92	-0.76	-0.60	-0.28	-0.12	0.12	0.44	1.0
y_i	1.84	1.96	2.21	2.45	2.94	3.18	4.9	5.2	5.7

At this point, it is more useful to express the normal equations in matrix-vector notation. Because this is how the computer works with the problem, it is instructive to examine the properties of these matrices. To get to matrix-vector form of the normal equations, employ a set of basis functions, $\{g_j(x)\}_{j=1}^{n}$, by (1) evaluating each basis function, $g_j(x)$, at each value of the explanatory variable, x_i, in the dataset, (2) forming the following matrix

$$\mathbf{A} = \begin{bmatrix} g_1(x_1) & g_2(x_1) & \dots & g_j(x_1) & \dots & g_n(x_1) \\ g_1(x_2) & g_2(x_2) & \dots & g_j(x_2) & \dots & g_n(x_2) \\ \vdots & & \ddots & & & \vdots \\ g_1(x_i) & g_2(x_i) & \dots & g_j(x_i) & \dots & g_n(x_i) \\ \vdots & & & & \ddots & \vdots \\ g_1(x_m) & g_2(x_m) & \dots & g_j(x_m) & \dots & g_n(x_m) \end{bmatrix} \tag{11}$$

and, (3) use this expression for $\mathbf{A} : (m+1) \times n$ to form $\mathbf{D} = \mathbf{A}^T\mathbf{A}$. Viewed in this light, the normal equations (in matrix-vector form) become

$$\mathbf{Dc} = \mathbf{A}^T\mathbf{y}$$

where $\mathbf{c}$ is a $n \times 1$ column vector corresponding to the coefficients on the n basis functions, and $\mathbf{y}$ is the $(m+1) \times 1$ column vector corresponding to the $m+1$ data points.

Based on this matrix representation, an important consideration for numerically solving the normal equations is the conditioning of $\mathbf{D}$. More concretely, we wish compare the degree of multicollinearity present in the $\mathbf{D}$ matrix under the standard basis and under the Chebyshev polynomial transformation. Computing the condition number for each of these matrices will give us a way to compare the two techniques.

Using Matlab, I constructed **D** for the present example using both the natural polynomial basis for $\mathbb{R}^{n-1}$, $\{1, x, x^2, ..., x^{n-1}\}$, and using the Chebyshev polynomial basis. Under the natural polynomial basis, the condition number for **D** is 570, whereas under the Chebyshev polynomial basis, the condition number for **D** is 6.19. This enormous difference in conditioning of the **D** matrix indicates a much greater stability from using Chebyshev polynomials to solve for the best fitting polynomial, as well as greater efficiency in obtaining solutions to the normal equations.[6]

5.3. **Gibbs Phenomena and Current Research.** There is a vast assortment of other applications of Chebyshev polynomials. These applications span from functional approximations to methods to solve partial differential equations. All of these applications perform poorly when the solution (or function to be approximated) has a discontinuity or is nearly discontinuous. Breakdowns in Chebyshev polynomial approximations around discontinuities are called Gibbs Oscillations (or the Gibbs Phenomenon, see [7]). In any application of Chebyshev polynomials, one must always be wary of discontinuities in the function (or solution) to be approximated.

Although Chebyshev methods do not handle discontinuities well, they appear to work optimally when the underlying function is continuous. In fact, current research on Chebyshev methods seeks to modify the approximations given by Chebyshev polynomials with spectral filters in order to remove Gibbs Oscillations. Sarra (2006) concludes his paper by saying, "Postprocessing methods to lessen the effects of the Gibbs oscillations are an active research area which would be an excellent topic for undergraduate research or as the topic of a Masters thesis." [7]

6. Conclusion

Pafnuty Chebyshev was a remarkable mathematician who made important contributions to many branches mathematics, including number theory, probability, the theory of integration, and, as is apparent from this paper, numerical analysis. Chebyshev accomplished a great deal with nearly every topic he studied. In no branch of thought is this more apparent than in numerical analysis. To say that Chebyshev polynomials are a useful tool for approximation is an understatement. Chebyshev polynomials are powerful, dynamic, and have changed the face of numerical analysis.

[6]The Matlab Code I employed to form the **D** matrix under different polynomial bases and to find the conditioning of each of these matrices is attached in the appendix. Matlab output from running the code is also attached.

References

[1] Boik, R. *A Pair of Primers: Primer on Matrix Analysis and Primer on Linear Statistical Models*. Department of Mathematical Sciences. Montana State University – Bozeman (August 13, 2007).

[2] Cheney, W. and D. Kincaid. 2003. *Numerical Mathematics and Computing*. Brooks Cole: New York. 5th Edition (July 25, 2003).

[3] Fischer, B. and R. Freund. 1989. "Chebyshev Polynomials are Not Always Optimal." Research Institute for Advanced Computer Science – NASA Ames Research Center. RIACS Technical Report 89.17.

[4] Iyanaga, S. and Kawada, Y (Eds.) 1977. *Cebysev (Tschebysheff) Polynomials*. Appendix A, Table 20.II in Encyclopedic Dictionary of Mathematics. Cambridge, MA: MIT Press, pp. 1478-1479.

[5] O'Connor, J.J. and E. F. Robertson. 2006. *Chebyshev Biography*. Turnbull WWW Server at the School of Mathematical and Computational Sciences, University of St. Andrews. Updated through August 2006. [URL: http://www-history.mcs.st-andrews.ac.uk/Biographies/Chebyshev.html]

[6] Sarra, Scott S. 2006. Chebyshev Interpolation: An Interactive Tour. *Journal of Online Mathematics and Its Applications*. Volume 6 (August 2006). Article ID: 1297.

[7] Weisstein, Eric. 2004. *Chebyshev Polynomial of the First Kind*. From Math-World–A Wolfram Web Resource. Last Modified: August 5, 2004. http://mathworld.wolfram.com/ChebyshevPolynomialoftheFirstKind.html. [Access Date: September 25, 2006].

[8] Wikipedia, the Free Encyclopedia. 2006. *Orthogonal Polynomials/Proofs*. Last updated: May 3, 2006. http://en.wikipedia.org/wiki/Orthogonal_polynomials/proofs.

Appendix A: Matlab Code and Output for the Example in Section 5

This appendix presents the Matlab code and output used to do the computations in the example in Section 5. First, here's the Matlab code:

```
% Enter Data

x=[1,1.1,1.3,1.5,1.9,2.1,2.4,2.8,3.5]
y=[1.84,1.96,2.21,2.45,2.94,3.18,4.9,5.2,5.7]

a=min(x); b=max(x);
xpoly=(2.*x-(a+b)).*(1/(b-a));

% Polynomial Basis D Matrix on [-1,1] %

fprintf(1,'---------------------------------------\n');
fprintf(1,'Working with Ordinary Polynomial Basis...\n');
fprintf(1,'---------------------------------------\n');

px1=xpoly.^0;
px2=xpoly.^1;
px3=xpoly.^2;
px4=xpoly.^3;
px5=xpoly.^4;
px6=xpoly.^5;

A=[px1',px2',px3',px4',px5'];
D=(A)'*A;
% Compute the Condition Number of D %

k=cond(D);

% Set the timer for the time it takes to find the polynomial solution %

t1=clock;

fprintf(1,'---------------------------------------\n');
fprintf(1,'The Condition Number for D is\n');
fprintf(1,'---------------------------------------\n');

k

fprintf(1,'---------------------------------------\n');
fprintf(1,'The Solution for the polynomial's coefficents is...\n');
fprintf(1,'---------------------------------------\n');

solution=inv(D)*A'*y';
poly(solution)

fprintf(1,'---------------------------------------\n');
fprintf(1,'Time Taken for Ordinary Basis\n');
fprintf(1,'---------------------------------------\n');
```

```
etime(clock, t1)

% Using Chebyshev Polynomials as a Basis  %

% Transform the data to the interval [-1,1]

xcheby=(2.*x-(a+b)).*(1/(b-a));

fprintf(1,'----------------------------------------\n');
fprintf(1,'Working with Chebyshev Polynomials...\n');
fprintf(1,'----------------------------------------\n');

% Obtain Chebyshev Polynomials up to order 5 %

cx1=xcheby.^0;
cx2=xcheby.^1;
cx3=2.*(xcheby.^2)-1;
cx4=4.*(xcheby.^3)-3.*xcheby;
cx5=8.*(xcheby.^4)-8.*(xcheby.^2)+1;
cx6=16.*(xcheby.^5)-20.*(xcheby.^3)+5.*(xcheby);

CA=[cx1',cx2',cx3',cx4',cx5'];
CD=(CA)'*CA;

% Compute the Condition Number for the Chebyshev D matrix %

k=cond(CD);

% Set the timer for the time it takes to find the polynomial solution %

t2=clock;

fprintf(1,'----------------------------------------\n');
fprintf(1,'The Condition Number for D is\n');
fprintf(1,'----------------------------------------\n');

k
solution=inv(CD)*(CA'*y');

fprintf(1,'----------------------------------------\n');
fprintf(1,'Time Taken for Chebyshev\n');
fprintf(1,'----------------------------------------\n');

etime(clock, t2)

% Recover the coefficients in the polynomial %

cheby=[1,0,0,0,0;0,1,0,0,0;-1,0,2,0,0;0,-3,0,4,0;1,0,-8,0,8];

fprintf(1,'----------------------------------------\n');
fprintf(1,'The Solution for the polynomial's coefficents is...\n');
```

```
fprintf(1,'----------------------------------------\n');

poly(solution'*cheby)
```

The Matlab code above yielded the following output.

```
a441project
x =
    1.0000    1.1000    1.3000    1.5000    1.9000
2.1000    2.4000    2.8000    3.5000

y =
    1.8400    1.9600    2.2100    2.4500    2.9400
3.1800    4.9000    5.2000    5.7000
----------------------------------------
Working with Ordinary Polynomial Basis...
----------------------------------------
----------------------------------------
The Condition Number for D is
----------------------------------------
k =
  507.2910
----------------------------------------
The Solution for the polynomial's coefficients is
----------------------------------------

ans =
    1.0000   -5.6803    0.6986   24.7772    1.5490  -13.6814
----------------------------------------
Time Taken for Ordinary Basis
----------------------------------------

ans =
    0.2010

----------------------------------------
Working with Chebyshev Polynomials...
----------------------------------------
----------------------------------------
The Condition Number for D is
----------------------------------------

k =
    6.1914

----------------------------------------
Time Taken for Chebyshev
----------------------------------------

ans =
    0.0900

----------------------------------------
```

```
The Solution for the polynomial's coefficients is
----------------------------------------
ans =
    1.0000   -5.6803    0.6986   24.7772    1.5490  -13.6814
```

Notice that not only did the Chebyshev method yield a better conditioned matrix, but the Chebyshev method was able to do the computations in less than half the computation time. For the present example (where there are only 9 data points and 6 coefficients), this does not make a big difference. For real world examples where computation time is an important factor, Chebyshev polynomials can both improve the numerical stability of the estimates and decrease the computational complexity of solving the normal equations.