Steven Pav spav@math.ucsd.edu

§10.2 Orthonormal Bases

Recall the general setting of least squares: we are given a set of data,

x	x_0	x_1	$\ldots$	x_n
y	y_0	y_1	$\ldots$	y_n

and some class of functions, $\mathcal{F}$, which is spanned by the basis vectors $\{g_j(x)\}_{j=0}^m$, *i.e.*,

$$\mathcal{F} = \left\{ f(x) = \sum_j c_j g_j(x) \mid c_j \in \mathrm{R},\ j = 0, 1, \ldots, m \right\}$$

We found that the least squares "best" function in $\mathcal{F}$ is $f = \sum_j c_j g_j$, where the c_j satisfy the *normal equations*:

$$\begin{bmatrix} d_{00} & d_{01} & d_{02} & \cdots & d_{0m} \\ d_{10} & d_{11} & d_{12} & \cdots & d_{1m} \\ d_{20} & d_{21} & d_{22} & \cdots & d_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{m0} & d_{m1} & d_{m2} & \cdots & d_{mm} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} e_0 \\ e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} \tag{1}$$

where

$$d_{ij} = \sum_{k=0}^n g_j(x_k) g_i(x_k), \quad e_i = \sum_{k=0}^n y_k g_i(x_k).$$

We already saw how poor choice of basis vectors can lead to numerical problems. Roughly speaking, if $g_i(x_k)$ is small for some i's and k's, then some d_{ij} can have a loss of precision when two small quantities are multiplied together and rounded to zero.

Poor choice of basis vectors can also lead to numerical problems in solution of the normal equations, which will be done by Gaussian Elimination.

We now consider the case where $\mathcal{F}$ is the class of polynomials of degree $\leq m$, and for simplicity we suppose that all $x_i \in [-1, 1]$. You might think to let $g_i(x) = x^i$. This would certainly give a basis for $\mathcal{F}$. However, this would be a poor choice of basis, as seen in Figure 1.

The short answer is that you should use, as a basis, the Chebyshev Polynomials of the first kind, *i.e.*, let $g_i(x) = T_i(x)$, where

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{i+1}(x) = 2xT_i(x) - T_{i-1}(x).$$

These polynomials are illustrated in Figure 2.

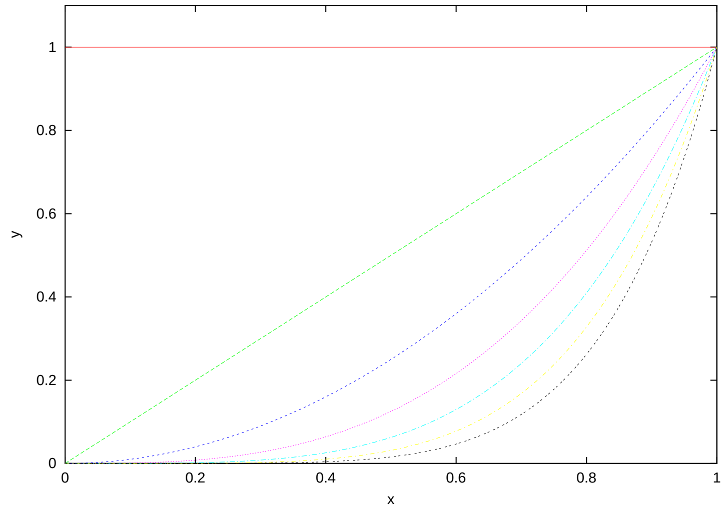

Figure 1: The polynomials x^i for $i = 0, 1, \ldots, 6$ are shown on $[0, 1]$. These polynomials make a bad basis because they look so much alike, essentially.

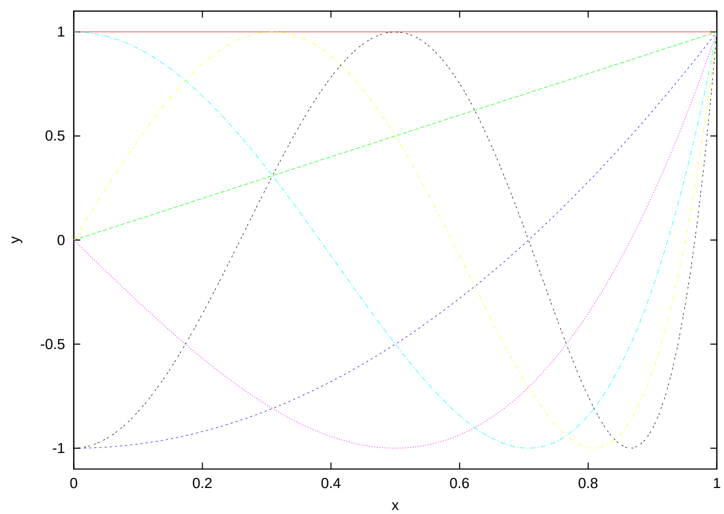

Figure 2: The Chebyshev polynomials $T_i(x)$ for $i = 0, 1, \ldots, 6$ are shown on $[0, 1]$. These polynomials make a better basis for least squares because they are orthogonal under some inner product. Basically, they do not look like each other.

Alternatives to Normal Equations

It turns out that the Normal Equations method isn't really so great. We consider other methods. First, we define A as the $n \times m$ matrix defined by the entries:

$$a_{ij} = g_j(x_i), \quad i = 0, 1, \ldots, n,\ j = 0, 1, \ldots, m.$$

That is

$$\mathsf{A} = \begin{bmatrix} g_0(x_0) & g_1(x_0) & g_2(x_0) & \cdots & g_m(x_0) \\ g_0(x_1) & g_1(x_1) & g_2(x_1) & \cdots & g_m(x_1) \\ g_0(x_2) & g_1(x_2) & g_2(x_2) & \cdots & g_m(x_2) \\ g_0(x_3) & g_1(x_3) & g_2(x_3) & \cdots & g_m(x_3) \\ g_0(x_4) & g_1(x_4) & g_2(x_4) & \cdots & g_m(x_4) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_0(x_n) & g_1(x_n) & g_2(x_n) & \cdots & g_m(x_n) \end{bmatrix}$$

We write it in this way since we are thinking of the case where $n \gg m$, so A is "tall." We now let $\boldsymbol{b}$ be defined by $b_i = y_i$.

After some inspection, we find that the Normal Equations can be written as:

$$\mathsf{A}^\top \mathsf{A}\,\boldsymbol{c} = \mathsf{A}^\top \boldsymbol{b}.$$

If $\boldsymbol{c}$ is a vector representing a function of $\mathcal{F}$ under the given basis, then the error, or residual, associated with this function is given by

$$\boldsymbol{b} - \mathsf{A}\,\boldsymbol{c}.$$

In our least squares theory we attempted to find that $\boldsymbol{c}$ that minimized

$$||\boldsymbol{b} - \mathsf{A}\,\boldsymbol{c}||_2^2 = (\boldsymbol{b} - \mathsf{A}\,\boldsymbol{c})^\top (\boldsymbol{b} - \mathsf{A}\,\boldsymbol{c})$$

We can see this as minimizing the Euclidian distance from $\boldsymbol{b}$ to $\mathsf{A}\,\boldsymbol{c}$. For this reason, we will have that the residual is orthogonal to the column space of A, that is we want

$$\mathsf{A}^\top (\boldsymbol{b} - \mathsf{A}\,\boldsymbol{c}) = \mathbf{0}.$$

This is just the normal equations. We could rewrite this, however, in the following form: find $\boldsymbol{c}$, $\boldsymbol{r}$ such that

$$\begin{bmatrix} \mathsf{I} & \mathsf{A} \\ \mathsf{A}^\top & \mathsf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ \mathbf{0} \end{bmatrix}$$

This is now a system of $n + m$ variables and unknowns, which can be solved by specialized means. This is known as the *augmented form.* We briefly mention that naïve Gaussian Elimination is not appropriate to solve the augmented form, as it is equivalent to using the normal equations method.