ORTHOGONAL POLYNOMIALS & MATRIX COMPUTATIONS

David F. Gleich

November 14, 2019

One of the most wonderful and surprising connections in the field of matrix computations is the elegant interplay between matrices and orthogonal polynomials. These relationships lead to magnificently simple insights into complicated methods. In this lecture we shall unshroud some of these connections. However, the field is so deep. Probably the best textbook on the topic was written by Purdue's own Walter Gautschi. It is tersely titled: "Orthogonal Polynomials."

The notes here arise from Saad, Iterative Methods, 2nd edition §6.72; Gutknecht, A General Framework for Recursions for Krylov Space Solvers, ETH SAM report 2005-09, ftp://ftp.sam.math.ethz.ch/pub/sam-reports/reports/reports2005/2005-09.pdf; and my own notes from working with Gene Golub.

1 WHAT ARE ORTHOGONAL POLYNOMIALS?

For the purposes of this lecture, a polynomial is a univariate function:[1]

$$p(t) = \sum_{i=0}^{n} p_i t^i.$$

For example:

$$p(t) = \frac{1}{2}t^2 - \frac{1}{2}$$
$$q(t) = \frac{5}{2}t^3 - \frac{3}{2}t.$$

The degree of a polynomial is the power of the largest term. In the generic polynomial definition, the degree is n. For the examples of p and q, the degrees are 2 and 3 respectively. A polynomial of degree 0 is a constant.

We call two polynomials orthogonal if:

$$\int_{-1}^{1} p(t)q(t)\,dt = 0.$$

This is a type of continuous analog of two vectors:

$$\mathbf{v}^T\mathbf{u} = \sum_{i=1}^{n} v_i u_i = 0.$$

EXAMPLE 1 *The two polynomials p and q given above are orthogonal.*

$$\begin{aligned}\int_{-1}^{1} p(t)q(t)\,dt &= \int_{-1}^{1} \frac{1}{2}(t^2-1)\frac{1}{2}(5t^3-3t)\,dt \\ &= \frac{1}{4}\int_{-1}^{1} 5t^5 - 5t^3 - 3t^3 + 3t\,dt \\ &= \frac{1}{4}\int_{-1}^{1} 5t^5 - 8t^3 + 3t\,dt.\end{aligned}$$

All of these terms are odd functions, and they are integrated over a symmetric region, hence the result is 0.

More generally, we can consider polynomials that are orthogonal with respect to

an arbitrary interval	$\int_a^b p(t)q(t)\,dt$
a weighted integral	$\int_a^b p(t)q(t)\,dw(t)$
a discrete weight	$\int_a^b p(t)q(t)\,dw(t) = \sum_{i=1}^{n} p(\lambda_i)q(\lambda_i)w_i$

If two polynomials are orthogonal with respect to an integral $\int_a^b dw(t)$, then we'll often call this the *measure* that they are orthogonal with respect to.

[1] Multivariate generalizations exist, but we won't need them.

SEQUENCES AND FAMILIES OF ORTHOGONAL POLYNOMIALS

We are often concerned with a sequence or family of orthogonal polynomials. We will index these by the degree of a polynomial, so that we have the sequence in order of increasing degree. Let $p_k(t)$ be the polynomial of degree k. By convention, we take $p_{-1}(t) = 0$ and $p_0(t) = c$ for some constant.

Thus, we have the following sequence of orthogonal polynomials that are orthogonal with respect to $\int_{-1}^{1} dt$:

$$\begin{aligned}
p_{-1}(t) &= 0\\
p_0(t) &= 1\\
p_1(t) &= t\\
p_2(t) &= \tfrac{1}{2}(3t^2 - 1)\\
p_3(t) &= \tfrac{1}{2}(5t^3 - t)\\
p_4(t) &= \tfrac{1}{8}(35t^4 - 30t^2 + 3)
\end{aligned}$$

This sequence is called the Legendre polynomials. Other popular families are:

Family	Measure
Legendre	$\int_{-1}^{1} dt$
Laguarre	$\int_{0}^{\infty} e^{-t}\, dt$
Hermite	$\int_{-\infty}^{\infty} e^{-t^2}\, dt$
Chebyshev (1st kind)	$\int_{-1}^{1} \frac{1}{\sqrt{1-t^2}}\, dt$
Chebyshev (2nd kind)	$\int_{-1}^{1} \sqrt{1-t^2}\, dt$
Jacobi	$\int_{-1}^{1} (1-t)^{\alpha}(1+t)^{\beta}\, dt$
Gegenbauer	$\int_{-1}^{1} (1-t^2)^{\alpha-1/2}\, dt$

Note that these can be scaled and shifted to arbitrary intervals $[a, b]$ too.

ORTHONORMAL POLYNOMIALS

Yes, there are orthonormal polynomials too! Checkout Gautschi's book for more on the relationship. If $\int_a^b dw(t)$ is the measure, the idea is that we need $\int_a^b p(t)^2\, dw(t) = 1$ in order to get orthonormal polynomials.

2 THE THREE TERM RECURRENCE & TRIDIAGONAL MATRICES

The following fact, which we will not prove, is profound:

THEOREM 2 *Any sequence of orthogonal polynomials of increasing degree satisfies a three-term recurrence and any three-term recurrence defines a sequence of orthogonal polynomials.*

EXAMPLE 3 *Consider the Legendre polynomials (described above). They satisfy:*

$$p_{k+1}(t) = \frac{2k+1}{k+1} t p_k(t) - \frac{k}{k+1} p_{k-1}(t).$$

For instance,

$$p_2(t) = \frac{2+1}{1+1} t \underbrace{t}_{p_1(t)} - \frac{1}{1+1} \underbrace{1}_{p_0(t)}.$$

The general form of the three-term recurrence is:[2]

$$p_{k+1}(t) = \mu_k\big(p_k(t) - \gamma_k t p_k(t)\big) + \eta_k p_{k-1}(t),$$

where the constants μ_k, γ_k, η_k may depend on k. What this recurrence means is that if we have any sequences of numbers μ_k, γ_k, η_k, then they give rise to a set of polynomials that is orthogonal with respect to *some measure*.

[2] In the notes I'm writing this from, μ_k is ρ_k, so beware of future typos.

FROM A THREE-TERM RECURRENCE TO A TRIDIAGONAL MATRIX

It's this three term recurrence that brings us back to matrix computations, and specificall tridiagonal matrices. Consider the Legendre family. Another way to write the recurrence is:

$$p_{-1}(t) = 0 \quad p_0(t) = 1 \quad p_1(t) = t \quad (k+1)p_{k+1}(t) = (2k+1)tp_k(t) - kp_{k-1}(t).$$

This recurrence forms a matrix:

$$\underbrace{\begin{bmatrix} 1 & & & & & \\ -t & 1 & & & & \\ 1 & -3t & 2 & & & \\ 0 & \ddots & \ddots & \ddots & & \\ & & k & -(2k+1)t & (k+1) & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}}_{\text{tridiagonal}\boldsymbol{T}} \underbrace{\begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \\ p_{k+1}(t) \\ \vdots \end{bmatrix}}_{\mathbf{p}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix}$$

Thus, for instance, we can *evaluate* a sequence of orthogonal polynomials at a point t by constructing $\boldsymbol{T}(t)$ and solving:

$$\boldsymbol{T}(t)\mathbf{p} = \mathbf{e}_1.$$

Consequently, *any* tridiagonal matrix corresponds to a set of orthogonal polynomials.

There's a Matlab demo of this in the `orthopolys.m` function!

There are additional relationships with a matrix called the Jacobi matrix:

$$\boldsymbol{J} = \begin{bmatrix} \alpha_1 & 1 & & \\ \beta_1 & \alpha_2 & 1 & \\ 0 & \ddots & \ddots & \ddots 0 \end{bmatrix}$$

but you'll need to read more about that in Gautschi's book.

3 POLYNOMIALS AND MATRICES

Note that if we have a univariate polynomial $p(t) = \sum_{i=0}^{n} p_i t^i$, then we can evaluate that polynomial with a *square* matrix argument:

$$p(\boldsymbol{A}) = \sum_{i=0}^{n} p_i \boldsymbol{A}^i.$$

4 POLYNOMIALS AND ITERATIVE METHODS

First, let us introduce the idea of using polynomials and iterative methods. This has been a homework or exam problem in the past, so it's worth understanding the details! Consider a Krylov subspace method. The kth iterate $\mathbf{x}_k$ is in $\mathbb{K}_k(\boldsymbol{A}, \mathbf{b}) = \operatorname{span}\{\mathbf{b}, \boldsymbol{A}\mathbf{b}, \ldots, \boldsymbol{A}^{k-1}\mathbf{b}\}$. This means that there is some polynomial such that:[3]

$$\mathbf{x}_k = s_{k-1}(\boldsymbol{A})\mathbf{b}.$$

Now, if $\mathbf{x}_k$ is determined by a polynomial $s_{k-1}(t)$ of degree $k-1$, this means the residual at the kth step is determined by a polynomial of degree k:

$$\mathbf{r}_k = \mathbf{b} - \boldsymbol{A}\mathbf{x}_k = \mathbf{b} - \boldsymbol{A}s_{k-1}(\boldsymbol{A})\mathbf{b} = \big(\boldsymbol{I} - \boldsymbol{A}s_{k-1}(\boldsymbol{A})\big)\mathbf{b}. \quad \text{Thus} \quad \mathbf{r}_k = p_k(\boldsymbol{A})\mathbf{b}$$

where the polynomial $p_k(t) = 1 - ts_{k-1}(t)$ has degree k.

We call these two polynomials:

$s_k(t)$ the solution polynomial
$p_k(t)$ the residual polynomial.

It turns out that the residual polynomial already must have some special structure! Note that $p_k(t)$ is defined to be equal to $1 - ts_{k-1}(t)$. Thus, $p_k(0) = 1$. So any residual polynomial must evaluate to 1 at the value $t = 0$. As a matrix statement, this means: $p_k(0)\mathbf{b} = \mathbf{b}$.

[3] This is the detail that you should work out!

5 ORTHOGONAL POLYNOMIALS AND ITERATIVE METHODS

Thus far, we haven't run into orthogonal polynomials yet. But let's design an iterative method with a fairly natural property using orthogonal polynomials.

Design goal We want the kth residual from the iterative method to be *orthogonal* to all previous residuals. Or more formally, $\mathbf{r}_k^T \mathbf{r}_j = 0$ for $j < k$. This goal is equivalent to the idea that our residual should always include *new* information at each step and should never include information we could have factored out.

Orthogonal polynomials will help us achieve this goal!

Let's state what we have:

$$\mathbf{r}_k = \mathbf{b} - \boldsymbol{A}\mathbf{x}_k = p_k(\boldsymbol{A})\mathbf{b}.$$

We want:

$$\mathbf{r}_k^T \mathbf{r}_j = \mathbf{b}^T p_k(\boldsymbol{A})^T p_j(\boldsymbol{A})\mathbf{b} = 0.$$

Thus, if we *create* an orthogonal polynomial $p_k(t)$ where $p_k(0) = 1$ and

$$\int p_k(t) p_j(t)\, dw(t) = (\mathbf{r}_k)^T \mathbf{r}_j,$$

we will implicitly create an iterative method where the residuals are orthogonal.[4]

The three term recurrence helps us do this!

Recall:

$$p_{k+1}(t) = \mu_k\big(p_k(t) - \gamma_k t p_k(t)\big) + \eta_k p_{k-1}(t).$$

So if we can determine μ_k, γ_k, η_k from our constraints, then we'll be able to figure out what the next residual polynomial is.

Let's enumerate our constraints:

$$\begin{aligned}&\text{(i)} p_{k+1}(0) = 1\\ &\text{(ii)}\mathbf{r}_{k+1} = p_{k+1}(\boldsymbol{A})\mathbf{b} = \mu_k\big(\mathbf{r}_k - \gamma_k \boldsymbol{A}\mathbf{r}_k\big) + \eta_k \mathbf{r}_{k-1}\\ &\text{(iii)}(\mathbf{r}_{k+1})^T \mathbf{r}_j = 0, j < k.\end{aligned}$$

Constraint (i) implies:

$$1 = p_{k+1}(0) = \mu_k + \eta_k \quad \Rightarrow \quad \eta_k = 1 - \mu_k.$$

Thus, we now have a revised constraint (ii):

$$\mathbf{r}_{k+1} = p_{k+1}(\boldsymbol{A})\mathbf{b} = \mu_k\big(\mathbf{r}_k - \gamma_k \boldsymbol{A}\mathbf{r}_k\big) + (1 - \mu_k)\mathbf{r}_{k-1}.$$

If we apply constraint (iii) with $j = k$, we have:

$$0 = (\mathbf{r}_k)^T \mathbf{r}_{k+1} = \mu_k\big((\mathbf{r}_k)^T \mathbf{r}_k - \gamma_k (\mathbf{r}_k)^T \boldsymbol{A}\mathbf{r}_k\big) + (1 - \mu_k)(\mathbf{r}_k)^T \mathbf{r}_{k-1}$$

or

$$\gamma_k = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_k^T \boldsymbol{A}\mathbf{r}_k}.$$

Finally, using constraint (iii) with $j = k - 1$, we can solve for[5]

$$\mu_k = \frac{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1}}{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1} + \gamma_k \mathbf{r}_{k-1}^T \boldsymbol{A}\mathbf{r}_k.}$$

This lets us compute $\mathbf{r}_{k+1}$

[4] Take a moment to understand what is going on here, as it's a key step. We are first noting that $\mathbf{r}_k$ can be expressed as a polynomial in $\boldsymbol{A}$. We are now saying, let's control that polynomial to achieve our goal! But we'll have to obey some constraints to make it work.

[5] Check for an error here... and this simplifies more, see Saad

GETTING THE SOLUTION POLYNOMIAL

Thus far, we have the residual polynomial $\mathbf{r}_k = p_k(\boldsymbol{A})\mathbf{b}$. We do, however, need to recreate the solution too! To do so, we note that

$$s_k(t) = \frac{1 - p_{k+1}(t)}{t} = \frac{1 - \mu_k(p_k(t) - \gamma_k t p_k(t)) - (1 - \mu_k)p_{k-1}(t)}{t}.$$

Now we add and subtract μ_k/t in order to rewrite this as:

$$\begin{aligned} s_k(t) &= \mu_k \left[\frac{1 - p_k(t)}{t} - \gamma_k p_k(t) \right] - (1 - \mu_k)\frac{1 - p_{k-1}(t)}{t} \\ &= \mu_k(s_{k-1}(t) - \gamma_k p_k(t)) - (1 - \mu_k)s_{k-2}(t). \end{aligned}$$

Hence, we have:

$$\mathbf{x}_{k+1} = \mu_k(\mathbf{x}_k - \gamma_k \mathbf{r}_k) - (1 - \mu_k)\mathbf{x}_{k-1}.$$

6 A POLYNOMIAL FORM OF CONJUGATE GRADIENT

In a small surprise, we've arrived at a new form of the conjugate gradient algorithm! The iterates generated by this method are mathematically equivalent to those generated by CG! The way to prove this is to show that the residuals constructed in CG automatically satisfy the same property and live in the same subspace, hence, they must be the same.