3 Eigenvalue Problems

In this chapter we look at some "iterative methods" for solving $Ax = \lambda x$, where A is a large (sparse) **symmetric** n by n matrix.

Eigenvalue problems, being nonlinear in nature, always require an iterative approach. The classical method (for a symmetric matrix) is, to use Householder transformations to obtain a tridiagonal matrix T (see § 3.3) similar to the original matrix A, i.e.,

$$T = Q^T AQ,$$

with all matrices involved of size n by n. The Householder matrices which give us the matrix Q are the tools to generate the tridiagonal matrix T, and therefore Q can be expected to be orthogonal to round-off precision. After tridiagonalization, the QR-algorithm is applied to the tridiagonal matrix T. This approach requires direct manipulation of the entire matrix A.

What we discuss here as an iterative method, is to generate a smaller tridiagonal matrix T_r (of size r by r, $r < n$), such that for an n by m matrix Q_r with orthonormal columns we have

$$AQ_r = Q_r T_r;$$

the eigenvalues of this T_r approximate the eigenvalues of A in some sense. This process can be accomplished without manipulating the matrix A, by forming matrix-vector products only. It is called the Lanczos Algorithm.

3.1 Rayleigh Quotient

The Rayleigh quotient is a key ingredient in the QR algorithm. It also features prominently in understanding Sturm sequences, which we will treat in § 3.3.

We assume that A is an n by n symmetric matrix, and denote by

$$\lambda_{\min} = \lambda_n = \min \lambda_j, \qquad \lambda_{\max} = \lambda_1 = \max \lambda_j,$$

the smallest and largest eigenvalue of A.

The Rayleigh Quotient (RQ) is a real-valued function defined on the set $\mathbb{R}^n - \{0\}$, namely

Definition 3.1 (Rayleigh Quotient, Residual) *The Rayleigh Quotient is*

$$\rho(x) = \frac{x^T A x}{x^T x}, \qquad x \neq 0. \tag{3.1}$$

$$r = r(x) = Ax - \rho(x)x \tag{3.2}$$

is called the residual of x.

The gradient of the function ρ can easily be computed as

$$\nabla\rho(x) = \frac{2}{x^T x}\,(Ax - \rho(x)x) = \frac{2}{x^T x} r(x). \tag{3.3}$$

Here are a few facts about the RQ:

Lemma 3.2 *(a)* $\rho(\alpha x) = \rho(x)$.

(b) $\lambda_{\min} \leq \rho(x) \leq \lambda_{\max}$.

(c) If x *is an eigenvector of* A*, then* $\rho(x) = \lambda$ *is the corresponding eigenvalue.*

(d) ρ *is stationary at the point* x *(*$\nabla\rho(x) = 0$*) if and only if* x *is an eigenvector of* A.

(e) Minimum residual property:

$$||(A - \rho(x)I)x|| \leq ||(A - \mu I)x||, \qquad \forall \mu \in \mathbb{C}.$$

(f) The residual $r = r(x)$ *is orthogonal to* x*, i.e.,* $r^T x = 0$.

The proof is left as an exercise.

Theorem 3.3 (Courant-Fisher) *If* $A = A^T$*, and if the eigenvalues of* A *are ordered*

$$\lambda_n \leq \lambda_{n-1} \leq \ldots \leq \lambda_2 \leq \lambda_1,$$

then

$$\lambda_j = \min_{\dim(S)=n-j+1} \max_{x \in S} \frac{x^T A x}{x^T x} \tag{3.4}$$

$$= \max_{\dim(S)=j} \min_{x \in S} \frac{x^T A x}{x^T x}. \tag{3.5}$$

For a proof see *Wilkinson, The Algebraic Eigenvalue Problem*, page 99.

Theorem 3.4 (Perturbation Theorem) *Let $x \neq 0$, $r = r(x)$, $\rho = \rho(x)$. Then (ρ, x) is an eigenpair for a matrix $A - M$ (i.e., $(A-M)x = \rho x$), with*

$$||M|| = \frac{||r||}{||x||}.$$

Proof. Verify that

$$M = \frac{1}{||x||^2}\left(xr^T + rx^T\right)$$

is the perturbation matrix. □

3.2 Lanczos Algorithm

We now return to the problem of tridiagonalizing the symmetric matrix A. Let us assume that q_1, q_2, ..., q_n are orthonormal vectors. We write our tridiagonal matrix T as

$$T = \begin{pmatrix} \alpha_1 & \beta_2 & 0 & \dots & & \\ \beta_2 & \alpha_2 & \beta_3 & 0 & \dots & \\ 0 & \ddots & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \ddots & \\ & & 0 & \beta_{n-1} & \alpha_{n-1} & \beta_n \\ & & & 0 & \beta_n & \alpha_n \end{pmatrix}.$$

Looking at the matrix relation $AQ = QT$, or

$$A \left(\begin{array}{c|c|c|c} | & | & | & | \\ q_1 & q_2 & q_3 & \dots \\ | & | & | & | \end{array}\right) =$$

$$\left(\begin{array}{c|c|c|c} | & | & | & | \\ q_1 & q_2 & q_3 & \dots \\ | & | & | & | \end{array}\right) \begin{pmatrix} \alpha_1 & \beta_2 & 0 & \dots & & \\ \beta_2 & \alpha_2 & \beta_3 & 0 & \dots & \\ 0 & \ddots & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \ddots & \\ & & 0 & \beta_{n-1} & \alpha_{n-1} & \beta_n \\ & & & 0 & \beta_n & \alpha_n \end{pmatrix},$$

and using the orthogonality of the q_j's, we obtain

$$\begin{array}{rcrl} Aq_1 & = & & \alpha_1 q_1 + \beta_2 q_2 \\ & \vdots & & \\ Aq_j & = & \beta_j q_{j-1}+ & \alpha_j q_j + \beta_{j+1} q_{j+1} \\ & \vdots & & \end{array}$$

with $\alpha_j = q_j^T A q_j = \rho(q_j)$. Therefore, if we have computed $q_1, \dots, q_k$, we can find q_{k+1} via

$$\beta_{k+1} q_{k+1} = (A - \alpha_k I) q_k - \beta_k q_{k-1},$$

which results in the so-called **Lanczos Process**:

Set $q_0 = 0$; $\beta_1 = 1$;
Choose q_1 arbitrarily, with $||q_1|| = 1$;
for $j = 1, 2, \dots, n-1$,
  $\alpha_j = q_j^T A q_j$
  $v = (A - \alpha_j I) q_j - \beta_j q_{j-1}$;
  $\beta_{j+1} = ||v||_2$;
  $q_{j+1} = v / \beta_{j+1}$.

This algorithm implements a fast calculation of the Krylov sequence, the orthonormal basis of the Krylov subspace

$$\mathcal{K}_k(A, q_1) = \text{span}(q_1, Aq_1, \dots, A^k q_1).$$

Note that $A = QTQ^T$ implies $A^k = QT^kQ^T$; naturally $q_1 = Qe_1$. Therefore,

$$\begin{array}{rcl} Aq_1 & = & AQe_1 = QTe_1 \\ A^2 q_1 & = & A^2 Q e_1 = QT^2 e_1 \\ \vdots & \vdots & \vdots \\ A^k q_1 & = & A^k Q e_1 = QT^k e_1, \end{array}$$

or, in matrix form,

$$\left(\begin{array}{c|c|c|c} | & | & | & | \\ q_1 & Aq_1 & A^2 q_1 & \dots \\ | & | & | & | \end{array} \right) = \underbrace{Q}_{\text{orthogonal}} \underbrace{\left(\begin{array}{c|c|c|c} | & | & | & | \\ e_1 & Te_1 & T^2 e_1 & \dots \\ | & | & | & | \end{array} \right)}_{\text{upper triangular}}.$$

Let us denote by Q_r the matrix containing the first r columns of the matrix Q, and by T_r the r by r principal submatrix of T, i.e.

$$T_r = \begin{pmatrix} \alpha_1 & \beta_2 & 0 & \dots & & & \\ \beta_2 & \alpha_2 & \beta_3 & 0 & \dots & & \\ 0 & \ddots & \ddots & \ddots & & & \\ \vdots & & \ddots & \ddots & \ddots & & \\ & & 0 & \beta_{r-1} & \alpha_{r-1} & \beta_r \\ & & & 0 & \beta_r & \alpha_r \end{pmatrix}. \tag{3.6}$$

Then, in general, the extreme eigenvalues of T_r are good approximations to the extreme eigenvalues of A, even for moderately small values of r. A one-sided bound follows easily from the Courant-Fisher Theorem 3.3:

$$\begin{aligned} \lambda_1(A) &= \max \frac{x^T A x}{x^T x} \geq \max_{x \in \mathcal{K}_{r-1} = \mathrm{span}(q_1, \dots, q_r)} \frac{x^T A x}{x^T x} \\ &= \max_{y \in \mathbb{R}^r} \frac{y^T Q_r^T A Q_r y}{y^T Q_r^T Q_r y} = \max_{y \in \mathbb{R}^r} \frac{y^T T_r y}{y^T y} = \lambda_1(T_r). \end{aligned}$$

A bound from the other side is given by

Theorem 3.5 *Let v_1 be the eigenvector of A for the largest eigenvalue λ_1, $Av_1 = \lambda_1 v_1$. If q_1 has a component in the direction of v_1, i.e., $q_1^T v_1 \neq 0$, then*

$$0 \leq \lambda_1(A) - \lambda_1(T_r) \leq \frac{\mathrm{const}}{\left[C_{r-1} \left(1 + 2 \left(\frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n} \right) \right) \right]^2},$$

where C_k denotes the k-th Chebyshev polynomial.

If we have a vector x_j in the Krylov subspace $\mathcal{K}_j = \mathrm{span}(q_1, Aq_1, \dots, A^j q_1)$, then the gradient

$$\nabla \rho(x_j) = \frac{2}{x_j^T x_j} (Ax_j - \rho(x_j) x_j)$$

is in $\mathcal{K}_{j+1}$. When we compute eigenvalues, we are looking to minimize or maximize $\rho(x)$, so by searching in the Krylov subspaces we include the directions of steepest descent/ascent. This is another reason why the Lanczos method gives good eigenvalue approximations.

3.3 Eigenvalues of a symmetric tridiagonal matrix

Before we explain how to compute eigenvalues of a symmetric tridiagonal matrix, we first recall a few properties of the symmetric eigenvalue problem, and then briefly sketch the "Householder" approach to tridiagonalization. The previous section shows how to obtain tridiagonal matrices via the Lanczos process.

Theorem 3.6 (Symmetric eigenvalue problem) *Let A be an n by n real symmetric matrix, $A^T = A$.*

(i) *All eigenvalues of A are real.*

(ii) *Eigenvectors corresponding to different eigenvalues are orthogonal.*

(iii) *The matrix A has a complete system of orthonormal eigenvectors v_1, v_2, ..., v_n ($v_j^T v_k = 0$ for $j \neq k$, and $v_j^T v_j = 1$).*

(iv) *The matrix V which has the orthonormal eigenvectors v_1, v_2, ..., v_n as its columns is an orthogonal matrix, $V^{-1} = V^T$. It diagonalizes the matrix A:*

$$A = V\Lambda V^T \quad \text{and} \quad \Lambda = V^T A V.$$

The matrix A is orthogonally *similar to the diagonal matrix Λ.*

Can we bring the matrix A into a form such that we can immediately find the eigenvalues (for example, triangular)? Remember, Gaussian elimination transforms the original system into one which can easily be solved. First we should note that the LU factorization is not helpful, as multiplication by an elementary matrix does not preserve the eigenvalues (the eigenvalues of U are not the same as the eigenvalues of A). We must perform a sequence of similarity transformations, hoping to generate as many zeros in our matrix as possible. Since A is symmetric, we should be able to restrict our attention to orthogonal similarity transformations, $A \to QAQ^T$, Q orthogonal. Let us try to create a triangular matrix using Householder matrices. We find a matrix Q_1, which zeros out the elements in the first column below the diagonal:

$$\underbrace{\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix}}_{A} \xrightarrow{Q_1} \underbrace{\begin{pmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{pmatrix}}_{Q_1 A} \xrightarrow{Q_1^T} \underbrace{\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix}}_{Q_1 A Q_1^T}. \tag{3.7}$$

Apparently we have been too ambitious. Multiplication by Q_1^T "destroys" the zeros in the first column. On reflection[2], it is no surprise that a direct attempt at triangularization must fail in general. Finding a finite algorithm to do this, would contradict the theorem by Galois – finding eigenvalues is equivalent to finding roots of polynomials! However, we can modify this idea to create a triangular matrix T which is orthogonally similar to A. i.e., $T = Q^T AQ$, Q orthogonal. The first step of this process consists of applying the matrix Q_1

$$Q_1 = \left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & H_1 & \\ 0 & & & \end{array}\right),$$

where H_1 is the Householder matrix which maps the vector consisting of the last three elements of the first column of A into a vector with zeros as its last two entries. The ensuing multiplication by Q_1^T from the right leaves the elements in the first column unchanged.

$$\underbrace{\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix}}_{A} \xrightarrow{Q_1} \underbrace{\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{pmatrix}}_{Q_1 A} \xrightarrow{Q_1^T} \underbrace{\begin{pmatrix} \times & \times & 0 & 0 \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{pmatrix}}_{Q_1 A Q_1^T}. \tag{3.8}$$

We also get zeros in the first row, as the matrix $Q_1 A Q_1^T$ must be symmetric. In the next step, we multiply by Q_2,

$$Q_2 = \left(\begin{array}{c|c} I_2 & 0 \\ \hline 0 & H_2 \end{array}\right), \quad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

[2] no pun intended

where H_2 maps the vector consisting of the last two elements of the second column of $Q_1 A Q_1^T$ into a vector with a zero in its last component. Multiplication by Q_2^T from the right leaves the first two columns of $Q_2 Q_1 A Q_1^T$ unchanged, and creates the symmetric tridiagonal matrix T which is similar to A. The eigenvalues of A are the same as the eigenvalues of T, and finding those of T is substantially cheaper.

Generally it is not a good idea to compute the eigenvalues via the characteristic polynomial. This is true, but we can find good algorithms using $\det(A - \lambda I) = 0$, as long as we **avoid computing the coefficients of the characteristic polynomial** $p(\lambda)$. There are some interesting results about tridiagonal (symmetric) matrices. So let us assume, we have transformed our original matrix A into the tridiagonal matrix T,

$$T = \begin{pmatrix} \alpha_1 & \beta_2 & 0 & \dots & & & \\ \beta_2 & \alpha_2 & \beta_3 & 0 & \dots & & \\ 0 & \beta_3 & \alpha_3 & \beta_4 & 0 & \dots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & 0 & \beta_{n-1} & \alpha_{n-1} & \beta_n \\ & & & & 0 & \beta_n & \alpha_n \end{pmatrix}.$$

Such a matrix is called *irreducible*, if all $\beta_j \neq 0$. If one of the β_j's is zero, our problem splits into two smaller ones. Therefore, we assume T to be irreducible. As before, we denote by T_r the r-th leading principal submatrix of T (see (3.6). Now we define the polynomials $p_r(\lambda)$ by

$$\begin{aligned} p_0(\lambda) &= 1, \\ p_r(\lambda) &= \det(T_r - \lambda I_r), \qquad \text{for } r = 1, \dots, n. \end{aligned}$$

I_r denotes the r by r identity matrix. So $p_r(\lambda)$ is the determinant of the r-th principal submatrix of $T - \lambda I$. If we expand the determinant of T_r with respect to its last row, we see that the polynomials p_r satisfy the following recurrence relation:

Lemma 3.7

$$\begin{aligned} p_0(\lambda) &= 1; \qquad p_1(\lambda) = \alpha_1 - \lambda; \\ p_k(\lambda) &= (\alpha_k - \lambda) p_{k-1}(\lambda) - \beta_k^2 p_{k-2}(\lambda), \qquad k > 1. \end{aligned} \tag{3.9}$$

This allows an efficient evaluation of $p(\lambda) = p_n(\lambda)$, the determinant of $T - \lambda I$; to find eigenvalues, we could use any nonlinear equation solver, for example, bisection, or the secant method. We can also use Newton's method, since the derivative $p'(\lambda) = p'_n(\lambda)$ is readily available from Lemma 3.7:

Lemma 3.8

$$\begin{aligned} p'_0(\lambda) &= 0; \qquad p'_1(\lambda) = -1; \\ p'_k(\lambda) &= -p_{k-1}(\lambda) + (\alpha_k - \lambda)p'_{k-1}(\lambda) - \beta_k^2 p'_{k-2}(\lambda). \end{aligned} \tag{3.10}$$

There is more! The sequence p_0, p_1, ..., p_n, is called a *Sturm sequence*. A simple property of this sequence, namely the number of sign changes in the sequence, tells us a lot about the eigenvalues of A.

Definition 3.9 *$N_r(\lambda)$ denotes the number of eigenvalues of T_r which are strictly less than λ. $N(\lambda) = N_n(\lambda)$ is the number of eigenvalues of T which are strictly less than λ.*

We have the following result:

Theorem 3.10 *$N_r(\lambda)$ equals the number of sign changes*[3] *in the Sturm sequence $p_0(\lambda)$, $p_1(\lambda)$, ..., $p_r(\lambda)$.*

It is easy to see that for $\lambda \to -\infty$, there will be no sign changes in the sequence, whereas for $\lambda \to \infty$ there will be n sign changes (alternating signs). If, for a moment we pretend that the β_j are all zero, then the result is obvious. Of course, this is no proof, but it makes the Theorem less mysterious. With this result we can easily and very cheaply find out how many eigenvalues there are in a given interval $[a, b)$ — the number is $N(b) - N(a)$. In many applications, one is not necessarily interested in the exact location of the eigenvalues, but maybe just bounds (for example, to exclude the possibility of resonance phenomena). For cases like these, and when small eigenvalues are needed to high precision, the theory presented here proves extremely useful.

[3] If one of the elements of the sequence is 0, we count it as $+0$ if the preceding element is positive, and -0 if the preceding element is negative, i.e., the appearance of a 0 is not counted as a sign change. It is not possible to have two consecutive zeros. So, $2, 0, 3$ has no sign change, but $-1, 0, 4$ counts as one sign change.

Example 3.11 *The exact eigenvalues of the matrix*

$$T = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

are $\pm\sqrt{2}$, *and* $1 \pm \sqrt{3}$. *We find the following Sturm sequences listed in Table 3.1.*

Table 3.1: Sturm sequences

λ	p_0	p_1	p_2	p_3	p_4	$N(\lambda)$
0	1	1	-1	-3	4	2
1	1	0	-1	-1	3	2
2	1	-1	1	1	-4	3

Thus we can conclude, there are two negative eigenvalues, two eigenvalues less than 1, no eigenvalues in $[0, 1]$, *three eigenvalues less than* 2, *end that there is exactly one eigenvalue in* $[1, 2]$. *Newton's method with the starting values* -2, 0, 2, *and* 3.5, *produces the sequences listed in Table 3.2*

Table 3.2: Newton sequences for eigenvalues of T in Example 3.11

-2.00000000000000	0	1.00000000000000	3.50000000000000
-1.66666666666667	-1.00000000000000	1.50000000000000	3.04983108108108
-1.48755186721992	-0.50000000000000	1.41406250000000	2.81435129636520
-1.42320458154758	-0.71875000000000	1.41421356376053	2.73963116628583
-1.41437549846909	-0.73193171277149	1.41421356237310	2.73212383593375
-1.41421361638248	-0.73205079748474	1.41421356237310	2.73205081444064
-1.41421356237310	-0.73205080756888	1.41421356237310	2.73205080756888
-1.41421356237310	-0.73205080756888	1.41421356237310	2.73205080756888
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$-\sqrt{2}$	$1-\sqrt{3}$	$\sqrt{2}$	$1+\sqrt{3}$

We now prove the Sturm Sequence Property:

Theorem 3.12 *For unreduced T_r the eigenvalues of T_{r-1} and T_r strictly interlace.*

Proof. First we show that the eigenvalues of B_{r-1} and B_r interlace, where B_{r-1} and B_r are principal submatrices of a symmetric matrix B. Denote the eigenvalues of B_r by κ_j, $j = 1, \ldots, r$, and the eigenvalues of B_{r-1} by ν_j, $j = 1, \ldots, r-1$. Recall the Courant-Fisher Theorem 3.3, and use the fact that

$$\left\{ \frac{x^T B_{r-1} x}{x^T x} \right\} = \left\{ \frac{x^T B_r x}{x^T x} : x_r = e_r^T x = 0 \right\}.$$

$$\begin{aligned}
\nu_k &= \min_{\{\dim S = r-k\}} \max_{\{x \in S \subset \mathbb{R}^{r-1}\}} \frac{x^T B_{r-1} x}{x^T x} \\
&= \min_{\{\dim S = r-k \;\&\; x_r = 0\}} \max_{\{x \in S \subset \mathbb{R}^{r}\}} \frac{x^T B_r x}{x^T x} \\
&\geq \min_{\{\dim S = r-k = (r+1)-(k+1)\}} \max_{\{x \in S \subset \mathbb{R}^{r}\}} \frac{x^T B_r x}{x^T x} = \kappa_{k+1};
\end{aligned}$$

on the other hand, we have

$$\max_{x \in S} \frac{x^T B_r x}{x^T x} \geq \max_{x \in S \;\&\; x_r = 0} \frac{x^T B_r x}{x^T x} = \max_{x \in S \subset \mathbb{R}^{r-1}} \frac{x^T B_{r-1} x}{x^T x},$$

and therefore $\kappa_k \geq \nu_k$. Combining the two results, we obtain the (weak) interlacing property

$$\kappa_{k+1} \leq \nu_k \leq \kappa_k.$$

For T_r and T_{r-1} we have the recurrence 3.7. Thus, if μ were a zero of both p_r and p_{r-1}, i.e., $p_r(\mu)$ and $p_{r-1}(\mu)$, the recursion and $\beta_r \neq 0$ would imply $p_{r-2}(\mu) = 0$, and by induction $p_0(\mu) = 0$, a contradiction! So for unreduced T, we have strong interlacing. □

As a corollary we get that all eigenvalues of an unreduced tridiagonal symmetric matrix are distinct.

To prove Theorem 3.10 we refer to the eigenvalues of T_r as $\mu_1, \ldots, \mu_r$, and to the ones of T_{r+1} as $\nu_1, \ldots, \nu_{r+1}$.

Proof. Let $s = N_r(\lambda)$, i.e.,

$$\mu_1 < \mu_2 < \ldots < \mu_s < \lambda \leq \mu_{s+1} < \ldots < \mu_r. \tag{3.11}$$

From the interlacing property we have

$$\nu_1 < \mu_1 < \nu_2 < \ldots < \nu_s < \mu_s,$$
$$\mu_{s+1} < \nu_{s+2} < \ldots < \nu_r < \mu_r < \nu_{r+1}.$$

Therefore, either s or $s+1$ of the ν_j's are less than λ. Clearly, we have

$$p_r(\lambda) = \prod_{j=1}^{r} (\mu_j - \lambda), \qquad p_{r+1}(\lambda) = \prod_{j=1}^{r+1} (\nu_j - \lambda). \tag{3.12}$$

Let us first assume that $\lambda \neq \nu_j$, $\lambda \neq \mu_j$. Then, if $\nu_{s+1} < \lambda$, the polynomial $p_{r+1}(\lambda)$ has one extra negative factor, and $N_{r+1}(\lambda) = N_r(\lambda)+1$; if $\nu_{s+1} > \lambda$, then there is no extra sign change. Now suppose $\nu_{s+1} = \lambda$. Then $p_{r+1}(\lambda) = 0$, and we defined this as "no sign change", i.e., there are s eigenvalues ν_j strictly smaller than λ. Finally, assume $\mu_{s+1} = \lambda$, i.e. $p_r(\lambda) = 0$. Then p_{r-1} has s zeros strictly less than λ (there is no sign change when going from p_{r-1} to p_r). We then have $\nu_{s+1} < \mu_{s+1} = \lambda$, i.e. there are $(s+1)$ zeros of p_{r+1}, and indeed, we will register an additional sign change. $\square$