Lecture 32 Chistov's Algorithm

Many important computational problems in algebra (such as the solution of polynomial equations) depend strongly on basic algorithms in linear algebra. In turn, many problems in linear algebra reduce to the computation of the rank of a matrix. This problem thus occupies a central position in computational algebra. *NC* algorithms for matrix rank were given by Ibarra, Moran, and Rosier in 1980 for matrices over the complex numbers [53] and over general fields in 1986 by Mulmuley [82]. We will devote a future lecture to this topic, but for now we lay the groundwork by showing how to calculate the characteristic polynomial of a matrix over an arbitrary field in *NC*.

The major limitation of Csanky's algorithm for computing the characteristic polynomial of a matrix is that it does not work in all fields, since it involves a division by k in (39). This won't be possible for example if the field is $\mathcal{Z}_p$ and k is a multiple of p. Berkowitz [11] and Chistov [18] gave the first deterministic *NC* algorithms for computing characteristic polynomials over arbitrary fields. Here we present Chistov's method [18].

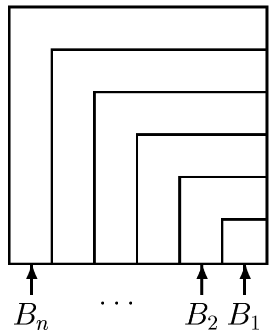
Recall that the characteristic polynomial of A, denoted $\chi_A(x)$, is defined by:

$$\begin{aligned}\chi_A(x) &= \det(xI - A) \\ &= x^n - s_1x^{n-1} + s_2x^{n-2} - \cdots \pm s_n \ .\end{aligned}$$

We will compute the polynomial that has the same coefficients, but in reverse order:

$$
\begin{aligned}
x^n \chi_A(\frac{1}{x}) &= 1 - s_1 x + s_2 x^2 + \cdots \pm s_n x^n \\
&= \det(I - xA) \; .
\end{aligned}
$$

Define $B = I - xA$ and let B_m denote the $m \times m$ submatrix in the lower right corner of B:



Let A_m be defined in the same way from A. Then $B_m = I_m - xA_m$. Define $\Delta_m = \det B_m$.

Cramer's rule gives a useful formula for the inverse of a matrix C in terms of determinants of its submatrices:

$$
C_{ij}^{-1} = (-1)^{i+j} \frac{\det \overline{C}_{ji}}{\det C}
$$

where $\overline{C}_{ji}$ denotes the submatrix obtained from C by removing the j^{th} row and i^{th} column. Applying Cramer's rule, we get

$$
(B_m^{-1})_{11} = \frac{\Delta_{m-1}}{\Delta_m} \; .
$$

But wait, this is all a bit suspicious, since B_m and Δ_m contain the indeterminate x. How can we invert a matrix with indeterminates? To make sense of this, we have to work in the field of *rational functions* over the base field k. This will let us divide by polynomials. The rational functions over k are the formal fractions

$$
k(x) = \left\{ \frac{p}{q} \;\middle|\; p, q \in k[x],\; q \neq 0 \right\} ,
$$

or more accurately, the equivalence classes of such fractions obtained by identifying p_1/q_1 and p_2/q_2 if $p_1 q_2 = p_2 q_1$. This construction is 100% analogous to the construction of the rational numbers from the integers.

Using the formal power series expansion of rational functions, the inverse of B_m can be expressed as an infinite formal sum

$$
B_m^{-1} = \sum_{i=0}^{\infty} x^i A_m^i \; . \tag{41}
$$

To convince yourself that this works, multiply (41) by $B_m = I_m - xA_m$. The expression (41) denotes a matrix of rational functions, because B_m is invertible as a linear map over the field $k(x)$: its determinant is $\Delta_m \neq 0$, as can be seen by evaluating at $x = 0$.

We can express $1/\Delta_n$, the determinant of B_n^{-1}, as a telescoping product like this:

$$\begin{aligned}
\frac{1}{\Delta_n} &= \frac{\Delta_{n-1}}{\Delta_n} \cdot \frac{\Delta_{n-2}}{\Delta_{n-1}} \cdots \frac{\Delta_0}{\Delta_1} \\
&= (B_n^{-1})_{11} \cdot (B_{n-1}^{-1})_{11} \cdots (B_1^{-1})_{11} \\
&= (\sum_{i=0}^{\infty} x^i A_n^i)_{11} \cdot (\sum_{i=0}^{\infty} x^i A_{n-1}^i)_{11} \cdots (\sum_{i=0}^{\infty} x^i A_1^i)_{11} && (42) \\
&= 1 - xH(x) \ , && (43)
\end{aligned}$$

where H is a humongous power series. The last step is justified by observing that the constant coefficients of all the factors in (42) are 1, therefore the constant coefficient of (43) is 1. Now recall that the polynomial we were originally looking for was Δ_n, which is the inverse of (43). We can therefore express Δ_n as a power series in terms of $H(x)$:

$$\begin{aligned}
\Delta_n &= \sum_{i=0}^{\infty} x^i H(x)^i \\
&= 1 - s_1 x + s_2 x^2 + \cdots \pm s_n x^n
\end{aligned}$$

and we know that the power series is a polynomial, so that all coefficients are zero after a certain point. Thus, despite all the infinite power series we have been using, all the terms after x^n vanish in the result. Therefore if we do all the calculations mod x^{n+1}, and take only the first $n+1$ terms of each series, we will still get the same answer.

This can be turned into a fast parallel algorithm, and since it involves no divisions, it will work in arbitrary fields.

32.1 The Characteristic Polynomial and Matrix Rank

The significance of the characteristic polynomial in matrix rank calculations is summed up in the following key lemma.

Lemma 32.1 *Let B be a square matrix over a field. If* $\operatorname{rank} B = \operatorname{rank} B^2$, *then* $\operatorname{rank} B = n - k$, *where x^k is the highest power of x that divides the characteristic polynomial $\chi_B(x)$.*

This lemma allows us to calculate the rank of a matrix by calculating its characteristic polynomial, provided its square has the same rank. A proper proof of this lemma would span a good portion of a first course in linear algebra, including Jordan canonical form and the Cayley-Hamilton Theorem, so it is a bit beyond our scope. Nevertheless, here it is in a nutshell.

When an $n \times n$ matrix B acts as a linear map on the vector space k^n, some vectors may be annihilated. These form a linear subspace called the kernel of B and denoted $\ker B$. The dimension of this subspace is $n - \operatorname{rank} B$. Vectors that are not annihilated by B get mapped around, and some may be mapped into the kernel, so that if the space is hit with B a second time, those vectors will be wiped out. The proviso $\operatorname{rank} B^2 = \operatorname{rank} B$ in Lemma 32.1 says that this does not happen. In other words, if a vector is ever going to be wiped out by some power of B, then it is already wiped out by B. For any B, the degree of the highest power of x that divides the characteristic polynomial of B is the dimension of the subspace of all vectors that ever get wiped out by some power of B. Thus if $\operatorname{rank} B^2 = \operatorname{rank} B$, then this subspace is just the kernel of B, and its dimension is $n - \operatorname{rank} B$.

The key property here is that the degree of the highest power of x that divides χ_B is the dimension of the subspace of all vectors that ever get wiped out by some power of B. Let's give this subspace a name:

$$\begin{aligned} E_0 &= \bigcup_{i=0}^{\infty} \ker B^i \\ &= \ker B^n . \end{aligned}$$

The last equation follows from the fact that the subspaces $\ker B^i$ are ordered by inclusion, $\ker B^i = n - \dim \operatorname{im} B^i$ ($\operatorname{im} B^i$ denotes the image of the whole space under the map B^i), and the image can only shrink in dimension n times before it disappears completely.

Another way of stating our key property is that $\dim E_0$ is the multiplicity of 0 as an eigenvalue of B. More generally, for each eigenvalue λ of B, we can define

$$\begin{aligned} E_\lambda &= \bigcup_{i=0}^{\infty} \ker (\lambda I - B)^i \\ &= \ker (\lambda I - B)^n . \end{aligned}$$

The subspace E_λ is called the *generalized eigenspace* of λ, and consists of all vectors of k^n that are annihilated by some power of the matrix $\lambda I - B$. The kernel of $\lambda I - B$ is called the *eigenspace* of λ.

Two nice things about the subspaces E_λ are that

(i) they are setwise invariant under the action of any matrix of the form $\mu I - B$; and

(ii) every vector can be represented *uniquely* as a sum of vectors, one from each generalized eigenspace.

Property (i) says that hitting the subspace E_λ repeatedly with the matrix $\lambda I - B$ does not move any vector outside of E_λ, but keeps shrinking it until it finally disappears; and if $\mu \neq \lambda$, then $\mu I - B$ is a bijection on E_λ. Property (ii) says that k^n is the *direct sum* of the subspaces E_λ; in symbols,

$$k^n \cong \bigoplus_\lambda E_\lambda$$

where $\cong$ denotes isomorphism of vector spaces and $\bigoplus$ denotes direct sum.

Now pick a new basis consisting of vectors in the subspaces E_λ. Under the change of basis, because of property (i), B becomes block diagonal with a block for each eigenvalue λ. (Judicious choice of these basis elements will even give us *Jordan canonical form*, with eigenvalues on the diagonal, 1's and 0's on the off-diagonal just above, and 0's elsewhere). The size of the block corresponding to λ is the dimension of E_λ. The change of basis is effected by a similarity transformation $B \mapsto U^{-1}BU$, which does not change the characteristic polynomial:

$$\begin{aligned} \det\,(xI - U^{-1}BU) &= \det U^{-1}(xI - B)U \\ &= \det U^{-1} \cdot \det\,(xI - B) \cdot \det U \\ &= \det\,(xI - B)\ . \end{aligned}$$

But the characteristic polynomial of a block diagonal matrix is the product of the characteristic polynomials of the blocks, which are $(x-\lambda)^{\dim E_\lambda}$. Thus

$$\chi_B(x) = \prod_\lambda (x-\lambda)^{\dim E_\lambda}\ . \tag{44}$$

If one of the eigenvalues is 0 (*i.e.*, if B has a nontrivial kernel), then $x^{\dim E_0}$ and no higher power of x will divide χ_B. This is what we wanted to show.

This conclusion also leads to an understanding of the *Cayley-Hamilton Theorem*: every matrix satisfies its characteristic equation. From (44) we get

$$\begin{aligned} \chi_B(B) &= \prod_\lambda (B - \lambda I)^{\dim E_\lambda} \\ &= \pm \prod_\lambda (\lambda I - B)^{\dim E_\lambda}\ . \end{aligned}$$

Applied to the whole space k^n, the factor

$$(\lambda I - B)^{\dim E_\lambda}$$

wipes out E_λ and fixes the other generalized eigenspaces setwise. Applying $\chi_B(B)$ to k^n applies these factors for each eigenvalue λ in succession, which successively wipe out all the E_λ, leaving nothing. Thus $\chi_B(B)$ is the zero matrix.