Dr H. Fawzi

Mathematical Tripos Part II: Michaelmas Term 2021

Numerical Analysis – Lecture 22

5.1 Simultaneous iteration

We assume in this section that $A \in \mathbb{R}^{n\times n}$ is a *symmetric* matrix, so that the eigenvectors associated to different eigenvalues, are orthogonal. Let $|\lambda_1| \geq \cdots \geq |\lambda_n|$ be the eigenvalues of A, and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ be eigenvectors. Consider the following algorithm, which generalizes the power method.

SIMULTANEOUS ITERATION – Let $X^{(0)} \in \mathbb{R}^{n\times p}$ has orthonormal columns
For $k = 0, 1, 2, \ldots$

- $Y = AX^{(k)}$
- $X^{(k+1)}R = \mathrm{qr}(Y)$ ($X^{(k+1)}$ has orthonormal columns, and R is upper triangular)

Revision 5.8 (QR factorization) *Recall that the QR factorization of a $n \times p$ matrix Y is $Y = QR$ where $Q \in \mathbb{R}^{n\times p}$ has orthonormal columns, and $R \in \mathbb{R}^{p\times p}$ is upper triangular. Such a factorization can be obtained by applying the Gram-Schmidt procedure on the columns of Y. Alternatively, it can be obtained using Householder reflections, or Givens rotations, see Numerical Analysis IB.*

The matrix $X^{(k)}$ produced by the algorithm above is nothing but the "Q" matrix in a QR factorization of $A^k X^{(0)}$. This can be easily seen by induction: it is clearly true for $k = 0$. Now assume that $A^k X^{(0)} = X^{(k)} R^{(k)}$ where $R^{(k)}$ is upper triangular. If we let $Y = AX^{(k)} = X^{(k+1)}R$ (the latter being a QR factorization, as per the algorithm above), then $A^{k+1}X^{(0)} = AX^{(k)}R^{(k)} = X^{(k+1)}RR^{(k)} = X^{(k+1)}R^{(k+1)}$ where $R^{(k+1)} = RR^{(k)}$ is upper triangular.

Relation with power method and inverse iteration It follows from the above, that the first column of $X^{(k)}$ is given by $X_1^{(k)} = A^k X_1^{(0)} / \|A^k X_1^{(0)}\|_2$, i.e., it corresponds to the power method starting from the vector $X_1^{(0)}$ (the first column of $X^{(0)}$).

Assume $p = n$. In this case it turns out that, remarkably, the last column of $X^{(k)}$, namely $X_n^{(k)}$, is the result of applying inverse iteration starting from the vector $X_n^{(0)}$ (the last column of $X^{(0)}$). Indeed if we invert the identity $A^k X^{(0)} = X^{(k)} R^{(k)}$ we get $(X^{(0)})^T A^{-k} = (R^{(k)})^{-1}(X^{(k)})^T$ [where we used the fact that $X^{(j)}$ are orthogonal matrices], and after transposing

$$A^{-k}X^{(0)} = X^{(k)}(R^{(k)})^{-T}.$$

Note that $(R^{(k)})^{-T}$ is *lower triangular*. This means that the last column of $A^{-k}X^{(0)}$ is a multiple of the last column of $X^{(k)}$, and so, by normalization, this means that

$$X_n^{(k)} = \frac{A^{-k}X_n^{(0)}}{\|A^{-k}X_n^{(0)}\|_2}.$$

This is precisely the result of applying inverse iteration (with shift $s = 0$) starting from $X_n^{(0)}$. This observation will be useful later when we introduce shifts in the QR iteration.

Convergence of simultaneous iteration The next theorem establishes convergence of the simultaneous iteration, and generalizes the statement for the power method. The theorem shows that colspan$(X^{(k)})$ converges to $\mathrm{span}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p)$ at the rate $(|\lambda_{p+1}|/|\lambda_p|)^k$, provided that the vectors in colspan$(X^{(0)})$ all have a nonzero component on $\mathrm{span}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p)$. To make the convergence statement precise, let $W = [\boldsymbol{w}_1|\ldots|\boldsymbol{w}_p] \in \mathbb{R}^{n\times p}$; and let $\bar{W} = [\boldsymbol{w}_{p+1}|\ldots|\boldsymbol{w}_n]$ which spans the orthogonal complement of colspan(W). We will show that $\bar{W}^T X^{(k)} \to 0$ at the rate $(|\lambda_{p+1}|/|\lambda_p|)^k$.

Theorem 5.9 *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues ordered in decreasing magnitude $|\lambda_1| \geq \cdots \geq |\lambda_n|$, and associated eigenvectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$. Assume that*

- $|\lambda_p| > |\lambda_{p+1}|$
- $X^{(0)} \in \mathbb{R}^{n\times p}$ *is such that* $W^T X^{(0)} \in \mathbb{R}^{p\times p}$ *is invertible, where* $W = [\boldsymbol{w}_1|\ldots|\boldsymbol{w}_p] \in \mathbb{R}^{n\times p}$.

Then $\|\bar{W}^T X^{(k)}\|_2 \leq c|\lambda_{p+1}/\lambda_p|^k$ where $\bar{W} = [\boldsymbol{w}_{p+1}|\ldots|\boldsymbol{w}_n]$, and $c > 0$ is a constant that depends on $X^{(0)}$ and $W, \bar{W}$. (Here $\|\cdot\|_2$ denotes the spectral norm.)

Proof. Let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ and $\bar{\Lambda} = \operatorname{diag}(\lambda_{p+1}, \ldots, \lambda_n)$, so that $AW = W\Lambda$, and $A\bar{W} = \bar{W}\bar{\Lambda}$.

We know from the earlier discussion that the matrix $X^{(k)}$ is obtained by orthonormalizing the columns of $A^k X^{(0)}$, more precisely, we have $A^k X^{(0)} = X^{(k)} R^{(k)}$, for some upper triangular $R^{(k)}$. This means that $X^{(k)} = A^k X^{(0)} (R^{(k)})^{-1}$, and thus:

$$\bar{W}^T X^{(k)} = \bar{W}^T A^k X^{(0)} (R^{(k)})^{-1} = \bar{\Lambda}^k \cdot (\bar{W}^T X^{(0)}) \cdot (R^{(k)})^{-1}. \tag{5.2}$$

where we used the fact that $\bar{W}^T A^k = \bar{\Lambda}^k \bar{W}^T$. In a very similar way we can write

$$W^T X^{(k)} = W^T A^k X^{(0)} (R^{(k)})^{-1} = \Lambda^k \cdot (W^T X^{(0)}) \cdot (R^{(k)})^{-1}. \tag{5.3}$$

By assumption, we know that $W^T X^{(0)} \in \mathbb{R}^{p\times p}$ is invertible. This allows us to eliminate $(R^{(k)})^{-1}$ in (5.2) using (5.3). Indeed we can write, using (5.3),

$$(R^{(k)})^{-1} = (W^T X^{(0)})^{-1} \cdot \Lambda^{-k} \cdot (W^T X^{(k)})$$

which, when plugging into (5.2) gives us

$$\bar{W}^T X^{(k)} = \bar{\Lambda}^k \cdot (\bar{W}^T X^{(0)}) \cdot (W^T X^{(0)})^{-1} \cdot \Lambda^{-k} \cdot (W^T X^{(k)}).$$

Now we can finish the proof:

$$\begin{aligned}\|\bar{W}^T X^{(k)}\|_2 &\leq \|\bar{\Lambda}^k\|_2 \cdot \|\bar{W}^T X^{(0)}\|_2 \cdot \|W^T X^{(0)})^{-1}\|_2 \cdot \|\Lambda^{-k}\|_2 \cdot \|W^T X^{(k)}\|_2 \\ &\leq c|\lambda_{p+1}|^k / |\lambda_p|^k\end{aligned}$$

where we used $\|\bar{\Lambda}^k\|_2 = |\lambda_{p+1}|^k$, $\|\Lambda^{-k}\|_2 = |\lambda_p|^{-k}$, and $\|W^T X^{(k)}\|_2 \leq 1$ since W and $X^{(k)}$ have orthonormal columns, and where $c = \|\bar{W}^T X^{(0)}\|_2 \cdot \|(W^T X^{(0)})^{-1}\|_2 > 0$. □

Consequence Assume that the eigenvalues all have distinct magnitudes, namely $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$, and consider applying simultaneous iteration with $p = n$. The theorem above shows that, for each $i = 1, \ldots, n-1$, columns 1 to i of $X^{(k)}$ will converge to $\operatorname{span}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_i)$. In particular, this implies that i'th column of $X^{(k)}$ will converge to $\pm\boldsymbol{w}_i$, so that $(X^{(k)})^T A X^{(k)} \to \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

QR iterations Given the above remark, it is useful to rewrite simultaneous iteration to keep track of the matrices $A^{(k)} = (X^{(k)})^T A X^{(k)}$. This gives the following, known as the basic QR iterations

BASIC QR ITERATION – Let $X^{(0)} \in \mathbb{R}^{n\times n}$ has orthonormal columns
Let $A^{(0)} = (X^{(0)})^T A X^{(0)}$. For $k = 0, 1, 2, \ldots$

- $QR = \operatorname{qr}(A^{(k)})$
- $A^{(k+1)} = RQ$

Note that the last line is equivalent to $A^{(k+1)} = Q^T A^{(k)} Q$: indeed, $Q^T A^{(k)} Q = Q^T (QR) Q = RQ$. It is not difficult to show by induction that the matrices $A^{(k)}$ produced by QR iteration, are the same as $(X^{(k)})^T A X^{(k)}$, where $X^{(k)}$ is produced by simultaneous iteration. Indeed, this is the case for $k = 0$. Now assume that $A^{(k)} = (X^{(k)})^T A X^{(k)}$ for some k. We know, from simultaneous iteration, that $X^{(k+1)}$ is obtained by performing a QR factorization of $AX^{(k)}$, i.e., $AX^{(k)} = X^{(k+1)} R^{(k+1)}$. Note that this automatically gives us a QR factorization of $A^{(k)} = (X^{(k)})^T A X^{(k)}$ since $X^{(k)}$ is orthogonal, namely: $A^{(k)} = QR$ where $Q = (X^{(k)})^T X^{(k+1)}$ and $R = R^{(k+1)}$. Now this allows us to show that $A^{(k+1)} = (X^{(k+1)})^T A X^{(k+1)}$: indeed

$$A^{(k+1)} = Q^T A^{(k)} Q = ((X^{(k+1)})^T X^{(k)})((X^{(k)})^T A X^{(k)})((X^{(k)})^T X^{(k+1)}) = (X^{(k+1)})^T A X^{(k+1)}.$$