Mathematical Tripos Part IB: Lent 2020

Numerical Analysis – Lecture 16[1]

Householder reflections Let $\boldsymbol{u} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}$. The $m \times m$ matrix $I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}$ is called a *Householder reflection*. Each such matrix is symmetric and orthogonal, since

$$\left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)^\top \left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right) = \left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)^2 = I - 4\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2} + 4\frac{\boldsymbol{u}(\boldsymbol{u}^\top\boldsymbol{u})\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^4} = I.$$

Householder reflections offer an alternative to Given rotations in the calculation of a QR factorization.

Householder algorithm Our goal is to multiply an $m \times n$ matrix A by a sequence of Householder reflections so that each product induces zeros under the diagonal in an entire column.

At the first step we seek a reflection that transforms the first column $\boldsymbol{a}_1$ of A to a multiple of $\boldsymbol{e}_1$. Since the Householder reflection is orthogonal (it preserves Euclidean norm) the latter has to be $\pm\|\boldsymbol{a}_1\|\boldsymbol{e}_1$ where we are free to choose the sign. The Householder reflection that does this operation is given by the choice of vector $\boldsymbol{u} = \boldsymbol{a}_1 - (\pm\|\boldsymbol{a}_1\|\boldsymbol{e}_1)$. For numerical stability the sign is usually chosen to be $-\mathrm{sign}(A_{11})$.

More generally, at the beginning of the k'th step of the algorithm, the columns 1 to $k-1$ have been processed and have zeros under their diagonal element. Our goal is to find a Householder reflection that will induce zeros under the diagonal element of the k'th column. To do so we use a block orthogonal matrix $\left[\begin{smallmatrix} I & 0 \\ 0 & H \end{smallmatrix}\right]$ where I is a $(k-1) \times (k-1)$ identity matrix, and H is a $(m-k+1) \times (m-k+1)$ Householder reflection associated with the choice $\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{a}}_k + \mathrm{sign}(A_{kk})\|\tilde{\boldsymbol{a}}_k\|\tilde{\boldsymbol{e}}_1$, where $\tilde{\boldsymbol{a}}_k$ is the vector of size $m-k+1$ consisting of the entries of A under the diagonal in the k'th column, and $\tilde{\boldsymbol{e}}_1$ is the vector of size $m-k+1$ with a 1 in the first position and zero elsewhere.

To summarize it is convenient to use the (Matlab-style) notation where $A_{k:m,j}$ indicates the vector of size $m-k+1$ obtained from rows $k,\ldots,m$ of column j of A. Then the algorithm can be written as follows:

Given $A \in \mathbb{R}^{m\times n}$ with $m \geq n$. For $k = 1$ to n:

- Let $\tilde{\boldsymbol{a}}_k = A_{k:m,k} \in \mathbb{R}^{m-k+1}$
- Let $\tilde{\boldsymbol{e}}_1$ be the vector of size $m-k+1$ with a 1 in the first position and zero elsewhere.
- Let $\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{a}}_k + \mathrm{sign}(A_{kk})\|\tilde{\boldsymbol{a}}_k\|\tilde{\boldsymbol{e}}_1$
- For each column $j = k,\ldots,n$ update $A_{k:m,j} = A_{k:m,j} - 2(\tilde{\boldsymbol{u}}^T A_{k:m,j})\tilde{\boldsymbol{u}}/\|\tilde{\boldsymbol{u}}\|^2$.

Example ($k = 3$, assuming the first two columns have already been processed)

$$A = \begin{bmatrix} 2 & 4 & 7 \\ 0 & 3 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix} \quad\rightarrow\quad \tilde{\boldsymbol{a}}_3 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix},\ \tilde{\boldsymbol{u}} = \begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix} \quad\rightarrow\quad \begin{bmatrix} 2 & 4 & 7 \\ 0 & 3 & -1 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Calculation of Q Like for the case of Givens algorithm, the matrix Q is not explicitly formed. To form Q explicitly we start with $\Omega = I$ initially and, for each step we replace Ω, by $\left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)\Omega = \Omega - \frac{2}{\|\boldsymbol{u}\|^2}\boldsymbol{u}(\boldsymbol{u}^\top\Omega)$ where $\boldsymbol{u} = \begin{bmatrix} \boldsymbol{0} \\ \tilde{\boldsymbol{u}} \end{bmatrix}$ is obtained from $\tilde{\boldsymbol{u}}$ by adding $k-1$ zeros above it[2]. However, if we require just the vector $\boldsymbol{c} = Q^\top\boldsymbol{b}$, say, rather than the matrix Q, then we set initially $\boldsymbol{c} = \boldsymbol{b}$ and in each stage replace $\boldsymbol{c}$ by $\left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)\boldsymbol{c} = \boldsymbol{c} - 2\frac{\boldsymbol{u}^\top\boldsymbol{c}}{\|\boldsymbol{u}\|^2}\boldsymbol{u}$.

[1] Corrections and suggestions to these notes should be emailed to `h.fawzi@damtp.cam.ac.uk`.

[2] Indeed, note that the reflection $I - 2\boldsymbol{u}\boldsymbol{u}^\top/\|\boldsymbol{u}\|^2$ is the same as the block orthogonal matrix $\left[\begin{smallmatrix} I & 0 \\ 0 & H \end{smallmatrix}\right]$ where H is the Householder reflection corresponding to $\tilde{\boldsymbol{u}}$.

Givens or Householder? If A is dense, it is in general more convenient to use Householder reflections. Givens rotations come into their own, however, when A has many leading zeros in its rows. E.g., if an $n \times n$ matrix A consists of zeros underneath the first subdiagonal, they can be 'rotated away' in just $n-1$ Givens rotations, at the cost of $\mathcal{O}(n^2)$ operations!

5.3 Linear least squares

Statement of the problem Suppose that an $m \times n$ matrix A and a vector $\boldsymbol{b} \in \mathbb{R}^m$ are given. The equation $A\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{x} \in \mathbb{R}^n$ is unknown, has in general no solution (if $m > n$) or an infinity of solutions (if $m < n$). Problems of this form occur frequently when we collect m observations (which, typically, are prone to measurement error) and wish to exploit them to form an n-variable linear model, where $n \ll m$. (In statistics, this is known as *linear regression*.) Bearing in mind the likely presence of errors in A and $\boldsymbol{b}$, we seek $\boldsymbol{x} \in \mathbb{R}^n$ that minimises the Euclidean length $\|A\boldsymbol{x} - \boldsymbol{b}\|$. This is the *least squares problem.*

Theorem $\boldsymbol{x} \in \mathbb{R}^n$ is a solution of the least squares problem iff $A^\top(A\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}$.

Proof. If $\boldsymbol{x}$ is a solution then it minimises

$$f(\boldsymbol{x}) := \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \langle A\boldsymbol{x} - \boldsymbol{b}, A\boldsymbol{x} - \boldsymbol{b}\rangle = \boldsymbol{x}^\top A^\top A\boldsymbol{x} - 2\boldsymbol{x}^\top A^\top \boldsymbol{b} + \boldsymbol{b}^\top \boldsymbol{b}.$$

Hence $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$. But $\frac{1}{2}\nabla f(\boldsymbol{x}) = A^\top A\boldsymbol{x} - A^\top \boldsymbol{b}$, hence $A^\top(A\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}$.
Conversely, suppose that $A^\top(A\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}$ and let $\boldsymbol{u} \in \mathbb{R}^n$. Hence, letting $\boldsymbol{y} = \boldsymbol{u} - \boldsymbol{x}$,

$$\begin{aligned}\|A\boldsymbol{u} - \boldsymbol{b}\|^2 &= \langle A\boldsymbol{x} + A\boldsymbol{y} - \boldsymbol{b}, A\boldsymbol{x} + A\boldsymbol{y} - \boldsymbol{b}\rangle = \langle A\boldsymbol{x} - \boldsymbol{b}, A\boldsymbol{x} - \boldsymbol{b}\rangle + 2\boldsymbol{y}^\top A^\top(A\boldsymbol{x} - \boldsymbol{b}) \\ &\quad + \langle A\boldsymbol{y}, A\boldsymbol{y}\rangle = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \|A\boldsymbol{y}\|^2 \geq \|A\boldsymbol{x} - \boldsymbol{b}\|^2\end{aligned}$$

and $\boldsymbol{x}$ is indeed optimal. □

Corollary Optimality of $\boldsymbol{x}$ ⇔ the vector $A\boldsymbol{x} - \boldsymbol{b}$ is orthogonal to all columns of A.

Normal equations One way of finding optimal $\boldsymbol{x}$ is by solving the $n \times n$ linear system $A^\top A\boldsymbol{x} = A^\top \boldsymbol{b}$; this is the method of *normal equations.* This approach is popular in many applications. However, there are three disadvantages. Firstly, $A^\top A$ might be singular, secondly sparse A might be replaced by a dense $A^\top A$ and, finally, forming $A^\top A$ might lead to loss of accuracy. Thus, suppose that our computer works in the IEEE arithmetic standard (≈ 15 significant digits) and let

$$A = \begin{bmatrix} 10^8 & -10^8 \\ 1 & 1 \end{bmatrix} \qquad \Longrightarrow \qquad A^\top A = \begin{bmatrix} 10^{16}+1 & -10^{16}+1 \\ -10^{16}+1 & 10^{16}+1 \end{bmatrix} \approx 10^{16} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Given $\boldsymbol{b} = [0, 2]^\top$ the solution of $A\boldsymbol{x} = \boldsymbol{b}$ is $[1, 1]^\top$, as can be easily found by Gaussian elimination. However, our computer 'believes' that $A^\top A$ is singular!

QR and least squares
Let A be an $m \times n$ matrix with $m \geq n$, and let $A = QR$ be a *reduced* QR factorization where Q is $m \times n$ has orthonormal columns and R is $n \times n$ upper triangular. We know that $\boldsymbol{x}$ is a solution to the least squares problem iff $A\boldsymbol{x} - \boldsymbol{b}$ is orthogonal to all columns of A. Since the columns of Q span the same space as the columns of A this is equivalent to saying that $Q^\top(A\boldsymbol{x} - \boldsymbol{b}) = 0$. Since the columns of Q form an orthonormal system we have[3] $Q^\top Q = I_n$, and so this leads to the equation $R\boldsymbol{x} = Q^\top \boldsymbol{b}$. The latter can be solved using backsubstitution.

[3]Note however that $QQ^\top$ is not equal to the identity matrix! (Q is a rectangular matrix here)