Mathematical Tripos Part IB: Lent 2010

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.

Deriving the first column of R 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. To start with, we seek a reflection that transforms the first nonzero column of A to a multiple of $\boldsymbol{e}_1$.
Let $\boldsymbol{a} \in \mathbb{R}^m$ be the first nonzero column of A. We wish to choose $\boldsymbol{u} \in \mathbb{R}^m$ s.t. the bottom $m-1$ entries of

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

vanish and, in addition, we normalise $\boldsymbol{u}$ so that $2\boldsymbol{u}^\top\boldsymbol{a} = \|\boldsymbol{u}\|^2$ (recall that $\boldsymbol{a} \neq \boldsymbol{0}$). Therefore $u_i = a_i$, $i = 2, \ldots, m$ and the normalisation implies that

$$2u_1a_1 + 2\sum_{i=2}^m a_i^2 = u_1^2 + \sum_{i=2}^m a_i^2 \quad \Rightarrow \quad u_1^2 - 2u_1a_1 + a_1^2 - \sum_{i=1}^m a_i^2 = 0 \quad \Rightarrow \quad u_1 = a_1 \pm \|\boldsymbol{a}\|.$$

It is usual to let the sign be the same as the sign of a_1, since otherwise $\|\boldsymbol{u}\| \ll 1$ might lead to a division by a tiny number, hence to numerical difficulties.

For large m we do not execute explicit matrix multiplication. Instead, to calculate

$$\left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)A = A - 2\frac{\boldsymbol{u}(\boldsymbol{u}^\top A)}{\|\boldsymbol{u}\|^2},$$

we first evaluate $\boldsymbol{w}^\top := \boldsymbol{u}^\top A$, subsequently forming $A - \frac{2}{\|\boldsymbol{u}\|^2}\boldsymbol{u}\boldsymbol{w}^\top$.

Subsequent columns of R Suppose that $\boldsymbol{a}$ is the first column of A that isn't compatible with standard form (previous columns have been, presumably, already dealt with by Householder reflections) and that the standard form requires to bring the $k+1, \ldots, m$ components to zero. Hence, nonzero elements in previous columns must be confined to the first $k-1$ rows and we want them to be unamended by the reflection. Thus, we let the first $k-1$ components of $\boldsymbol{u}$ be zero and choose $u_k = a_k \pm \left(\sum_{i=k}^m a_i^2\right)^{1/2}$ and $u_i = a_i$, $i = k+1, \ldots, m$.

The Householder method We process columns of A in sequence, in each stage premultiplying a current A by the requisite Householder reflection. The end result is an upper triangular matrix R in its standard form.

Example

$$A = \begin{bmatrix} 2 & 4 & 7 \\ 0 & 3 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{u} = \begin{bmatrix} 0 \\ 0 \\ 5 \\ 1 \\ -2 \end{bmatrix} \quad \Rightarrow \quad \left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)A = \begin{bmatrix} 2 & 4 & 7 \\ 0 & 3 & -1 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Calculation of Q If the matrix Q is required in an explicit form, set $\Omega = I$ initially and, for each successive reflection, replace Ω by $\left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)\Omega = \Omega - \frac{2}{\|\boldsymbol{u}\|^2}\boldsymbol{u}(\boldsymbol{u}^\top\Omega)$. As in the case of

[1]Corrections and suggestions to these notes should be emailed to `A.Iserles@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartIB/Handouts.html`.

Givens rotations, by the end of the computation, $Q = \Omega^\top$. 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 $\left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}\right)\boldsymbol{c} = \boldsymbol{c} - 2\frac{\boldsymbol{u}^\top\boldsymbol{c}}{\|\boldsymbol{u}\|^2}\boldsymbol{u}$.

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. $\square$

Corollary Optimality of $\boldsymbol{x} \Leftrightarrow$ 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 $A^\top A\boldsymbol{x} = A^\top\boldsymbol{b}$ – 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

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

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
Lemma Let A be any $m \times n$ matrix and let $\boldsymbol{b} \in \mathbb{R}^m$. The vector $\boldsymbol{x} \in \mathbb{R}^n$ minimises $\|A\boldsymbol{x} - \boldsymbol{b}\|$ iff it minimises $\|\Omega A\boldsymbol{x} - \Omega\boldsymbol{b}\|$ for an arbitrary $m \times m$ orthogonal matrix Ω.

Proof. Given an arbitrary vector $\boldsymbol{v} \in \mathbb{R}^m$, we have

$$\|\Omega\boldsymbol{v}\|^2 = \boldsymbol{v}^\top\Omega^\top\Omega\boldsymbol{v} = \boldsymbol{v}^\top\boldsymbol{v} = \|\boldsymbol{v}\|^2.$$

In particular, $\|\Omega A\boldsymbol{x} - \Omega\boldsymbol{b}\| = \|A\boldsymbol{x} - \boldsymbol{b}\|$. $\square$

Method of solution Suppose that $A = QR$, a QR factorization with R in a *standard form.* Because of the lemma, letting $\Omega := Q^\top$, we have $\|A\boldsymbol{x} - \boldsymbol{b}\| = \|Q^\top(A\boldsymbol{x} - \boldsymbol{b})\| = \|R\boldsymbol{x} - Q^\top\boldsymbol{b}\|$, therefore we seek $\boldsymbol{x} \in \mathbb{R}^n$ that minimises $\|R\boldsymbol{x} - Q^\top\boldsymbol{b}\|$.
In general ($m > n$) many rows of R consist of zeros. Suppose for simplicity that $\operatorname{rank} R = \operatorname{rank} A = n$. Then the bottom $m-n$ rows of R are zero. We find $\boldsymbol{x}$ by solving the (nonsingular) linear system given by the first n equations of $R\boldsymbol{x} = Q^\top\boldsymbol{b}$. Similar (but more complicated) algorithm applies when $\operatorname{rank} R \leq n-1$. Note that we don't require Q explicitly, just to evaluate $Q^\top\boldsymbol{b}$.