

Mathematical Tripos Part IB: Lent 2010

Numerical Analysis – Lecture 16¹

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

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

Householder reflections offer an alternative to Givens 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 \mathbf{e}_1 .

Let $\mathbf{a} \in \mathbb{R}^m$ be the first nonzero column of A . We wish to choose $\mathbf{u} \in \mathbb{R}^m$ s.t. the bottom $m - 1$ entries of

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

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

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

It is usual to let the sign be the same as the sign of a_1 , since otherwise $\|\mathbf{u}\| \ll 1$ 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

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

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

Subsequent columns of R Suppose that \mathbf{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 $k + 1, \dots, m$ 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 \mathbf{u} be zero and choose $u_k = a_k \pm (\sum_{i=k}^m a_i^2)^{1/2}$ and $u_i = a_i$, $i = k + 1, \dots, m$.

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

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

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 $\left(I - 2\frac{\mathbf{u}\mathbf{u}^\top}{\|\mathbf{u}\|^2}\right)\Omega = \Omega - \frac{2}{\|\mathbf{u}\|^2}\mathbf{u}(\mathbf{u}^\top\Omega)$. As in the case of

¹Corrections 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 $\mathbf{c} = Q^\top \mathbf{b}$, say, rather than the matrix Q , then we set initially $\mathbf{c} = \mathbf{b}$ and in each stage replace \mathbf{c} by $\left(I - 2 \frac{\mathbf{u}\mathbf{u}^\top}{\|\mathbf{u}\|^2}\right) \mathbf{c} = \mathbf{c} - 2 \frac{\mathbf{u}^\top \mathbf{c}}{\|\mathbf{u}\|^2} \mathbf{u}$.

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 $\mathbf{b} \in \mathbb{R}^m$ are given. The equation $A\mathbf{x} = \mathbf{b}$, where $\mathbf{x} \in \mathbb{R}^n$ 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 \mathbf{b} , we seek $\mathbf{x} \in \mathbb{R}^n$ that minimises the Euclidean length $\|A\mathbf{x} - \mathbf{b}\|$. This is the *least squares problem*.

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

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

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

Hence $\nabla f(\mathbf{x}) = \mathbf{0}$. But $\frac{1}{2}\nabla f(\mathbf{x}) = A^\top A\mathbf{x} - A^\top \mathbf{b}$, hence $A^\top(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$.

Conversely, suppose that $A^\top(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$ and let $\mathbf{u} \in \mathbb{R}^n$. Hence, letting $\mathbf{y} = \mathbf{u} - \mathbf{x}$,

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

and \mathbf{x} is indeed optimal. □

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

Normal equations One way of finding optimal \mathbf{x} is by solving the $n \times n$ linear system $A^\top A\mathbf{x} = A^\top \mathbf{b}$ – 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

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

Given $\mathbf{b} = [0, 2]^\top$ the solution of $A\mathbf{x} = \mathbf{b}$ 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 $\mathbf{b} \in \mathbb{R}^m$. The vector $\mathbf{x} \in \mathbb{R}^n$ minimises $\|A\mathbf{x} - \mathbf{b}\|$ iff it minimises $\|\Omega A\mathbf{x} - \Omega \mathbf{b}\|$ for an arbitrary $m \times m$ orthogonal matrix Ω .

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

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

In particular, $\|\Omega A\mathbf{x} - \Omega \mathbf{b}\| = \|A\mathbf{x} - \mathbf{b}\|$. □

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 $\|A\mathbf{x} - \mathbf{b}\| = \|Q^\top(A\mathbf{x} - \mathbf{b})\| = \|R\mathbf{x} - Q^\top \mathbf{b}\|$, therefore we seek $\mathbf{x} \in \mathbb{R}^n$ that minimises $\|R\mathbf{x} - Q^\top \mathbf{b}\|$.

In general ($m > n$) many rows of R consist of zeros. Suppose for simplicity that $\text{rank } R = \text{rank } A = n$. Then the bottom $m - n$ rows of R are zero. We find \mathbf{x} by solving the (nonsingular) linear system given by the first n equations of $R\mathbf{x} = Q^\top \mathbf{b}$. Similar (but more complicated) algorithm applies when $\text{rank } R \leq n - 1$. Note that we don’t require Q explicitly, just to evaluate $Q^\top \mathbf{b}$.