Normal equations

$$2\left(\mathbf{A}^t\mathbf{A}\mathbf{c}-\mathbf{A}^t\mathbf{y}\right)=0 \quad \text{or} \quad \mathbf{A}^t\mathbf{A}\mathbf{c}=\mathbf{A}^t\mathbf{y}$$

- The system is called the “Normal equations”
- Can solve least squares problems using these
- For **A** size $m \times n$ and **c** of size n and **y** of size m what are the dimensions of the normal equations?
  - $n \times n$
- Have converted it to a regular system that we know how to solve
- Solve via LU decomposition
- Solution should be accurate if the matrix $\mathbf{A}^t\mathbf{A}$ is well conditioned

More on Normal Equations

- Normal equations are only important theoretically
- Gives us a way to think about least squares.
- In practice least squares solved via a different process
  - QR decomposition
- Why?
  - Somewhat expensive as we have to form $A^t A$
  - involves matrix multiplication and then solution
  - More importantly it is poorly conditioned
  - $\text{cond}(A^tA) = (\text{cond}(A))^2$
- Would like a method whose errors are closer to the condition number of A

Look at the fitting matrix in more detail

- Suppose we want to solve via least squares

$$\mathbf{Ac=y}$$

  - **A** is a $m \times n$ matrix with $m>n$
- One way to solve was via LU decomposition of normal equations
  - Poor condition numbers and so not recommended
  - Requires matrix-matrix multiplication which is expensive
- Instead
  - Look for methods that can directly operate on **A** to get the solution
  - Recall in LU we did a set of transformations to **A** and the r.h.s. to find **c**
  - Today we will look at the QR algorithm

Key Ideas

- Column space of a matrix: the vector space formed by the collection of column vectors in a matrix

$$\begin{bmatrix} 1 & 8 & 13 & 12 \\ 14 & 11 & 2 & 7 \\ 4 & 5 & 16 & 9 \\ 15 & 10 & 3 & 6 \end{bmatrix}$$

- Every matrix vector product results in a vector formed by linear combination of vectors in the column space
- A $m \times n$ rectangular matrix $\mathbf{A}$ takes n vectors into m vectors
- Let the least squares problem be $\mathbf{Ac}=\mathbf{f}$
- Let the solution which minimizes the residual be $\mathbf{c}_*$
- Then $\mathbf{c}_*$ creates on matrix vector product a rhs $\mathbf{f}_*$ that is in the column space of $\mathbf{A}$
- We want that $\mathbf{c}_*$ minimizes $\mathbf{r}=||\mathbf{f} - \mathbf{f}_*||$

Column Space of A

- Not all *m* vectors **f** will be reachable even if we supply arbitrary *n* vectors **c**
- *Range* of A: the part of the space of *m* vectors that are reachable

$$Range(\mathrm{A}) = \{\mathbf{y} \in R^m : \mathbf{y}=\mathbf{A}\mathbf{x} \text{ for some } \mathbf{x} \in R^n\}$$

  – The range of A contains all those vectors that can be made up using the columns of **A**

  – *Rank*(**A**) is the dimension of the range of **A**

  – Null space of **A**: those vectors x, for which Ax is zero

$$\mathrm{Null}(\mathbf{A}) = \{\mathbf{x} \in R^n : \mathbf{A}\mathbf{x}=0\}$$

$$\mathrm{Dim}(\mathrm{Null}(\mathbf{A}))+\mathrm{Rank}(\mathbf{A})=n$$

- Key idea: We want to minimize the error in the part that can be reached

QR decomposition

- Suppose we can write

$$\mathbf{A}=\mathbf{Q}'\mathbf{R}'$$

  – $\mathbf{Q}'$ is an orthonormal matrix of dimension $m \times m$
  – Columns of $\mathbf{Q}'$ form a basis in the m dimensional space
  – However $\mathbf{A}$ has only n vectors
  – $\mathbf{R}'$ is a $m \times n$ matrix that can be written as $\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}$

  $\mathbf{R}$ is triangular $n \times n$ and $\mathbf{0}$ is a matrix of zeroes of size $m\text{-}n \times n$

  $\mathbf{Q}'$ can be partitioned as $[\mathbf{Q}\ \mathbf{Q}^{\sim}]$ with $\mathbf{Q}$ containing n orthonormal columns m spanning column space of $\mathbf{A}$

  $\mathbf{Q}^{\sim}$ $m\text{-}n$ orthonormal columns from unreachable part

- If $\mathbf{Ax}=\mathbf{b}$ then $(\mathbf{Q}'\ \mathbf{R}')\mathbf{x}=\mathbf{b}$ or $\mathbf{Q}'(\mathbf{R}'\mathbf{x})=\mathbf{b}$ or $\mathbf{Q}'\mathbf{y}=\mathbf{b}$
  – So if $\mathbf{b}$ is in range($\mathbf{A}$), it is also in range($\mathbf{Q}'$)
  – Similarly if $\mathbf{Q}'\mathbf{y}=\mathbf{b}$; then $\mathbf{b}=\mathbf{Ax}$ with $\mathbf{x}=\mathbf{R}^{-1}\mathbf{y}$
  – Columns of $\mathbf{Q}$ form an orthonormal basis for range($\mathbf{A}$)

Orthogonal Matrices

- Orthogonal matrices are square matrices that have their columns orthonormal to each other
  - dot product of different column vectors is zero, while of the same column is one
  - Denoted $\mathbf{Q}$
  - Most trivial orthogonal matrix is the identity matrix
  - For an orthonormal matrix
  - $\mathbf{Q}^t\mathbf{Q}=\mathbf{I}$
  - So $\mathbf{Q}^{-1}=\mathbf{Q}^T$

generalization: a complex matrix is *Hermitian* iff $\mathbf{Q}^{-1}=\mathbf{Q}^H$ where superscript ^{H} denotes complex conjugate transpose

Orthogonal matrix facts

- Suppose **Q** is an orthonormal matrix
- Then for any vector **r** the Euclidean norm is preserved in an orthonormal transformation
- Proof

$$||\mathbf{Q}r||^2=(Qr)^t\,(Qr)= r^t\,Q^t\,Q\,r = r^t\,(Q^t\,Q)\,r =r^t\,r =||r||^2$$

- If Q is an orthonormal matrix so is the extended matrix Q_e

$$Q_e = \begin{bmatrix} I & 0 \\ 0 & Q \end{bmatrix}$$

- Easy to show from definition that

$$Q_e^t\,Q_e = I$$

Solving least squares with QR

- $A=Q'R'$
- Let $r= y-Ac$ $b=Q'^t y$
- Goal of least squares find the c that minimizes squared error (residue)
- Partition b in to two pieces
  - b_1 of dimension n
  - b_2 of dimension m-n
  - $||r||^2 = || y- Ac||^2 = ||y - Q' R' c||^2$
  - Length is not changed by multiplication with orthogonal matrix
  - So $||r||^2 =||Q'^t r||^2 =||Q'^t [y - Q' R' c]||^2$

$$=||b_1 - R c ||^2 + ||b_2 - 0c||^2$$

So no matter what c is the second term remains unchanged

If we minimize $||r||^2$ the best we can do is minimize first term

Solving LS via QR

- How do we minimize $\|c_1 - R\, x\|^2$
  - If R is full rank set solve Rx=c then we have done the best we can
  - (if R is rank deficient solve in least squares sense)
  - Recall R is triangular so this equation can be easily solved
- Algorithm
  - Compute QR factorization of A=Q'R'
  - Form $c_1=Q^t\, b$
  - Solve $Rx=c_1$
  - If R is full rank and $Q^{\sim}$ is available then the norm of the residual is $\|Q^{\sim t}\, b\|$. Else $r = b - A\, x$.

Computing the factorization

- QR is useful … so how do we factorize a matrix A?
- In LU we computed a upper triangular matrix by computing adding multiples of other rows so that elements below a given column were zeroed out
- The multipliers were stored in L which gave us A=LU
- Here we want to zero out entries below the diagonal but do it with orthogonal matrices
- Two strategies
- Zero out a column at a time using a matrix Q_1 so that Q^t_1 A gives us all entries below a certain one in a column as zero
  - Householder transformations
  - Result $Q^t_n \ldots Q^t_2 Q^t_1 A = R$ or $A = Q_1 \ldots Q_{n-1} Q_n R = Q R$
- Zero out one specific entry of a column at a time
  - Givens rotations
- Product of orthogonal matrices is orthogonal

To compute QR

- Perform a sequence of orthogonal transformations that zero out elements
- Orthogonal transformations can be rotations or reflections or combinations
- Givens Rotation:
- Givens matrix has elements

$$G = \begin{bmatrix} c & s \\ s & -c \end{bmatrix}$$

- $c^2+s^2=1$
- How do we use a rotation to zero out an element?
- Let z= $[z_1\ z_2]^t$
- We want

$$Gz = \begin{bmatrix} cz_1 + sz_2 \\ sz_1 - cz_2 \end{bmatrix} = xe_1$$

- Eliminate z_2

$$(c^2 + s^2)z_1 = cx\,, \qquad c = z_1/x\,.$$

- Similarly we get $s=z_2/x$, and $z_1^2+z_2^2 = x^2$

Givens QR

- To apply idea to larger matrix, embed the Givens matrix in identity matrix. We will use the notation G_{ij} to denote an $n \times n$ identity matrix with its ith and jth rows modified to include the Givens rotation: for example, if $n = 6$, then

$$G_{25} = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{c} & 0 & 0 & \mathbf{s} & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & \mathbf{s} & 0 & 0 & -\mathbf{c} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix},$$

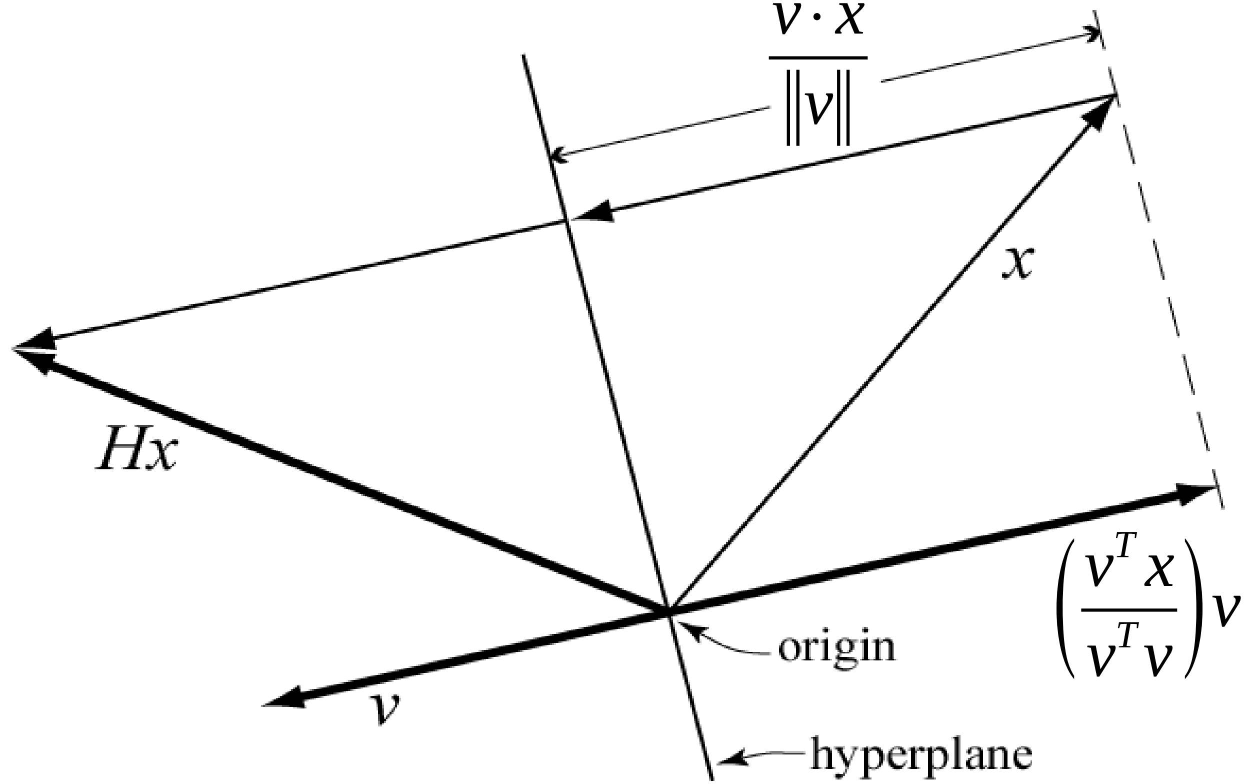
and multiplication of a vector by this matrix leaves all but rows 2 and 5 of the vector unchanged.

- Algorithm

```
for i=1, ...,n
    for j=i+1, ..., m
        Find Givens matrix G_ij to zero out j,i element of A
        using the the value at position (i,i)
        A=G_ij A
    end
end
```

Householder Geometry

- ***Hx* is *x* reflected through the hyperplane perpendicular to *v*** ($p : p^{\mathrm{T}}v=0$)



Householder Transformations

The *Householder transformation* determined by vector v is:

$$H = I - 2\frac{vv^T}{v^Tv}$$

outer product, n×n matrix

inner product, scalar

To apply it to a vector x, compute:

$$Hx = \left(I - 2\frac{vv^T}{v^Tv} \right) x = x - 2\frac{v(v^Tx)}{v^Tv}$$

$$\boxed{Hx = x - \left(2\frac{v^Tx}{v^Tv} \right) v}$$

scalar

Householder Properties

- H is symmetric, since

$$H^T = \left(I - 2\frac{vv^T}{v^Tv} \right)^T = I^T - 2\frac{(vv^T)^T}{v^Tv} = I - 2\frac{v^{T^T}v^T}{v^Tv} = H$$

- H is orthogonal, since

$$H^TH = HH = \left(I - 2\frac{vv^T}{v^Tv} \right)\left(I - 2\frac{vv^T}{v^Tv} \right)$$

$$= I - 4\frac{vv^T}{v^Tv} + 4\frac{v(v^Tv)v^T}{(v^Tv)^2} = I - 4\frac{vv^T}{v^Tv} + 4\frac{vv^T}{v^Tv} = I$$

and $H^TH = I$ implies $H^T = H^{-1}$

Householder to Zero Matrix Elements

We'll use Householder transformations to zero subdiagonal elements of a matrix.

Given any vector a, find the v that determines an H such that,

$$Ha = \alpha e_1 = \alpha[1, 0, 0, \ldots, 0]^T$$

Now solve for v:

$$Ha = a - \left(2\frac{v^T a}{v^T v}\right)v = a - \mu v = \alpha e_1$$

where μ is parenthesized scalar, related to length of v

$$\Rightarrow \; v = (a - \alpha e_1)/\mu$$

We're free to choose $\mu = 1$, since $\|v\|$ does not affect H

Choosing the Vector v

So $v = a - \alpha e_1$ for some scalar α.

But $\|Ha\|_2 = \|a\|_2$

(prove this by expanding $\|Ha\|_2^2 = (Ha)^T Ha$)

and $\|Ha\|_2 = |\alpha|$ by design, so $\alpha = \pm\|a\|_2$

(either sign will work).

To avoid $v \approx 0$ we choose $\alpha = -\text{sign}(a_1)\|a\|_2$,

so $v = a + \text{sign}(a_1)\|a\|_2 e_1$ is our answer.

Applying Householder Transforms

- Don't compute Hx explicitly, that costs $3n^2$ flops.
- Instead use the formula given previously,

$$Hx = x - \left(2 \frac{v^T x}{v^T v} \right) v$$

which costs $4n$ flops (if you pre-compute $v^T v$ or pre-normalize $v^T v = 2$).

- Typically, when using Householder transformations, you never compute the matrix H; it's only used in derivation and analysis.

QR Decomposition

- Householder transformations are a good way to zero out subdiagonal elements of a matrix.
- A is decomposed:

$$Q^T A = \begin{bmatrix} R \\ 0 \end{bmatrix} \quad \text{or} \quad QQ^T A = A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

- where $Q^T = H_n \ldots H_2 H_1$ is the orthogonal product of Householders and R is upper triangular.
- Overdetermined system $Ax=b$ is transformed into the easy-to-solve

$$\begin{bmatrix} R \\ 0 \end{bmatrix} x = Q^T b$$

- Cost: 4mn+4(m-1)(n-1)+…4(m-n) =O(mn^2)