Computational Methods

CMSC/AMSC/MAPL 460

Least squares method: QR

Ramani Duraiswami,

Dept. of Computer Science

Several slides in this lecture are adapted from Profs. O'Leary UMD and Heckbert, CMU

Look at the fitting matrix in more detail

- Suppose we want to solve via least squares

$$Ax=b$$

  - 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
  - Require matrix-matrix multiplication
- Today: Other matrix decompositions that are stable and less expensive
- Here A is a matrix that takes n vectors into m vectors
- Not all m vectors will be reachable even if we supply arbitrary n vectors
  - Range of A: the part of the space of m vectors that are reachable

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

  - 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

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

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

Orthogonal Matrices

- Orthogonal matrices are square matrices that have their columns orthonormal to each other
  - dot product of dfferent column vectors is zero, while of the same column is one
  - Denoted Q
  - Most trivial orthogonal matrix is the identity matrix
  - $Q^tQ=I$

So $Q^{-1}=Q^T$

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

QR decomposition

- Suppose we can write

$$A=Q'R'$$

  - Q' is an orthogonal matrix of dimension $m \times m$
  - R' is a $m \times n$ matrix that can be written as $\begin{bmatrix} R \\ 0 \end{bmatrix}$

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

  Q' can also be partitioned as [Q Q~] with Q containing n orthogonal columns and Q~ m-n orthogonal columns

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

Properties of QR

- Let $A^t z = 0$
  - z is in the nullspace of A^t
  - $(Q'R')^t z = R'^t Q'^t z = 0$
  - So $R^t y = 0$ for $y = Q^t z$
  - If R is full rank this means y has to be the zero vector
  - So $Q^t z = 0$
  - So z must be composed of the elements from $Q^\sim$
  - So the columns of $Q^\sim$ form an orthonormal basis for nullspace(A^t)

Orthogonal matrix facts

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

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

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

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

- Easy to show from definition that

$$Q_e^t\, Q_e = \mathrm{I}$$

Solving least squares with QR

- $A=Q'R'$
- Let $r= b-Ax$ $c=Q'^t b$
- Goal of least squares find the x that minimizes squared error (residue)
- Partition c in to two pieces
  - c_1 of dimension n
  - c_2 of dimension $m-n$
  - $||r||^2 = || b- Ax||^2 = ||b -Q' R' x||^2$
  - Length is not changed by multiplication with orthogonal matrix
  - So $||r||^2 =||Q'^t r||^2 =||Q'^t [b -Q' R' x]||^2 = ||c - \begin{bmatrix} R \\ 0 \end{bmatrix} x||^2$

    $=||c_1 - R x ||^2 + ||c_2 - 0x||^2$

So no matter what x 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
  - 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

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

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

- Algorithm

for $i=1, \ldots, n$

  for $j=i+1, \ldots, m$

    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 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:

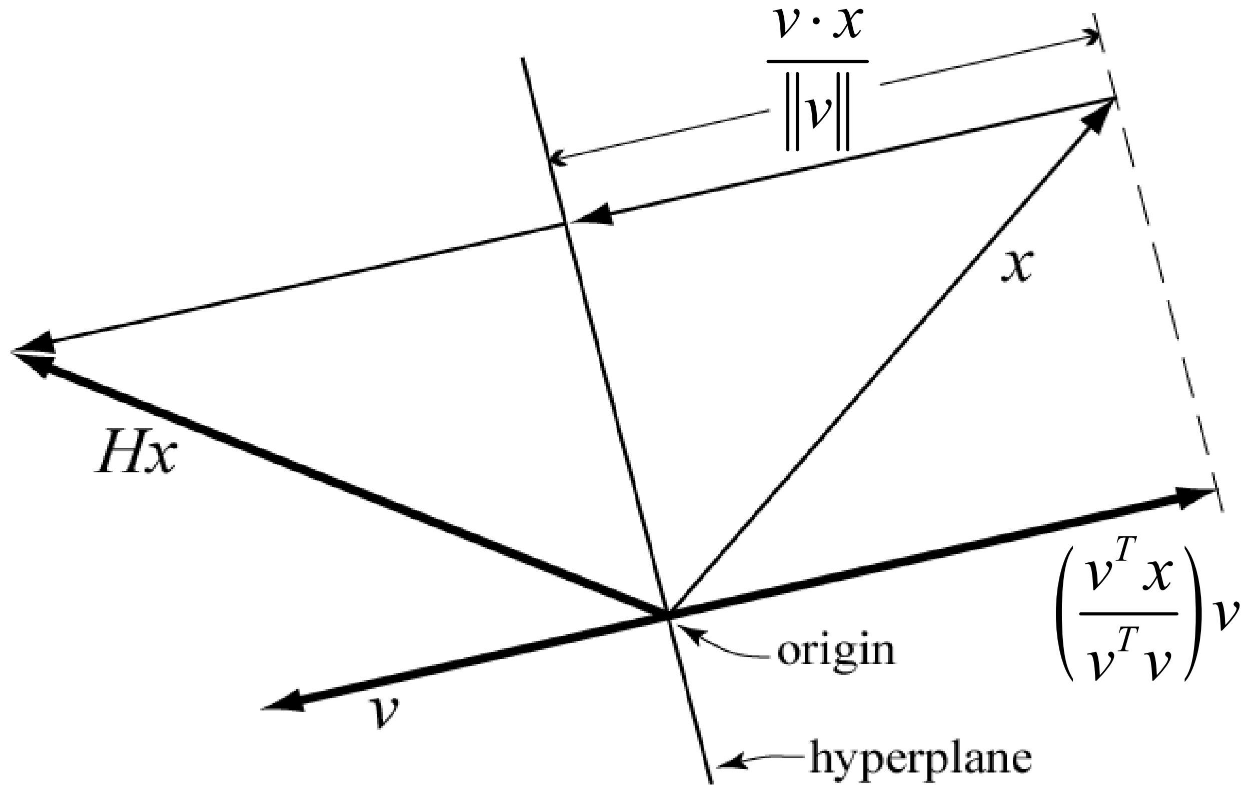
$$Hx = \left(I - 2\frac{vv^T}{v^Tv} \right) x = x - 2\frac{v(v^Tx)}{v^Tv}$$

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

scalar

Householder Geometry

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



Householder Properties

- H is symmetric, since

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

- H is orthogonal, since

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

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

and $H^T H = I$ 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 $v^{\mathrm{T}}v$ or pre-normalize $v^{\mathrm{T}}v=2$).
- 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 $Q^{\mathrm{T}}=H_n \ldots H_2 H_1$ is the orthogonal product of Householders and R is upper triangular.
- Overdetermined system $Ax=b$ is transformed into the easy-to-solve

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

SVD and Pseudo-Inverse

- $\mathbf{Ax}=\mathbf{b}$ $\mathbf{A}$ is $m \times n$, $\mathbf{x}$ is $n \times 1$ and $\mathbf{b}$ is $m \times 1$.
- $\mathbf{A}=\mathbf{USV}^t$ where $\mathbf{U}$ is m×m, $\mathbf{S}$ is m×n and $\mathbf{V}$ is $n \times n$
- $\mathbf{USV}^t\, \mathbf{x}=\mathbf{b}$. So $\mathbf{SV}^t\, \mathbf{x}=\mathbf{U}^t\mathbf{b}$
- If $\mathbf{A}$ has rank r, then r singular values are significant

  $\mathbf{V}^t\mathbf{x}= \mathrm{diag}(\sigma_1^{-1},\ldots,\sigma_r^{-1}, 0, \ldots, 0)\mathbf{U}^t\mathbf{b}$

  $\mathbf{x}= \mathbf{V}\mathrm{diag}(\sigma_1^{-1},\ldots,\sigma_r^{-1}, 0, \ldots, 0)\mathbf{U}^t\mathbf{b}$

$$\mathbf{x}_r = \sum_{i=1}^{r} \frac{\mathbf{u}_i^t \mathbf{b}}{\sigma_i} \mathbf{v}_i \qquad \sigma_r > \varepsilon, \quad \sigma_{r+1} \le \varepsilon$$

- Pseudoinverse $\mathbf{A}^{+}=\mathbf{V}\, \mathrm{diag}(\sigma_1^{-1},\ldots,\sigma_r^{-1}, 0, \ldots, 0)\, \mathbf{U}^t$
  - $\mathbf{A}^{+}$ is a $n \times m$ matrix.
  - If rank $(\mathbf{A}) = n$ then $\mathbf{A}^{+}=(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}$
  - If $\mathbf{A}$ is square $\mathbf{A}^{+}=\mathbf{A}^{-1}$