Computational Methods
CMSC/AMSC/MAPL 460

Least squares methods

Ramani Duraiswami,
Dept. of Computer Science
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 $Q_1^t A$ gives us all entries below a certain one in a column as zero
    - Householder transformations
    - Result $Q_1^t\ldots Q_2^t Q_1^t A = R$ 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
  - \( c^2 + s^2 = 1 \)
  - How do we use a rotation to zero out an element?
    - Let \( z = [z_1 \ z_2]^t \)
    - We want to eliminate \( z_2 \)
    - Eliminate \( z_2 \)
      \[
      (c^2 + s^2)z_1 = cx, \quad c = z_1/x.
      \]
    - 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 $i$th and $j$th rows modified to include the Givens rotation: for example, if $n = 6$, then

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

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

- Algorithm

\[
\text{for } i=1, \ldots, n \\
\quad \text{for } j=i+1, \ldots, m \\
\quad\quad \text{Find Givens matrix } G_{ij}\text{ to zero out } j,i \text{ element of } A \\
\quad\quad \text{using the the value at position } (i,i) \\
\quad\quad A = G_{ij}A \\
\quad \text{end} \\
\text{end}
\]
Goal: reflect axial vector through a (hyper)-plane

- Axial vector is any vector through the origin. Let it be \( \mathbf{u} \)

- Goal reflect it through a plane
- Let \( \mathbf{a} \) be the unit vector normal to the plane
- \( \mathbf{u} \cdot \mathbf{a} = \|\mathbf{u}\| \cos \theta \)
- To get reflected vector in the plane
  Subtract twice the component of \( \mathbf{u} \) along \( \mathbf{a} \)
  \( \mathbf{v} = \mathbf{u} - 2 \|\mathbf{u}\| \mathbf{a} \|\mathbf{u}\| \cos \theta \)
  \( \cos \theta = \frac{(\mathbf{a} \cdot \mathbf{u})}{\|\mathbf{u}\|} \)
Householder transform

• Achieve this reflection via multiplication by an orthogonal matrix

\[ v = Qu = u - 2 \frac{a \|u\|}{a \cdot u} \cos \theta \]

\[ = u - 2 \frac{a \|u\| (a \cdot u)}{(\|u\|^2)} \]

\[ = (I - 2 \frac{aa^t}{a^t a})u \]

• What if \( a \) is not a unit vector?

\( (I - 2 \frac{(aa^t)}{(a^t a)})u \)

• \( Q = I - 2 \frac{aa^t}{(a^t a)} \)
Householder Transformations

The *Householder transformation* determined by vector \( v \) is:

\[
H = I - 2 \frac{vv^T}{v^Tv} \quad \text{outer product, n×n matrix}
\]

\[
Hx = x - 2 \frac{v(v^Tx)}{v^Tv} \quad \text{inner product, scalar}
\]

To apply it to a vector \( x \), compute:

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

\[
Hx = x - \left( 2 \frac{v^Tx}{v^Tv} \right) v \quad \text{scalar}
\]
Householder Geometry

- $Hx$ is $x$ reflected through the hyperplane perpendicular to $v$ ($p : p^Tv=0$)
Householder Properties

• $H$ is symmetric, since

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

• $H$ is orthogonal, since

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

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

and $H^T H = I$ 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,

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

Now solve for $v$:

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

where $\mu$ 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 $\nu$

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

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

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

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

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

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

• 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:
  \[
  Q^T A = \begin{bmatrix} R \\ 0 \end{bmatrix} \quad \text{or} \quad Q Q^T A = A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}
  \]
  
- 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
  \[
  \begin{bmatrix} R \\ 0 \end{bmatrix} x = Q^T b
  \]
Other Norms

• Here we fit using the “least-squares” or $L_2$ norm
• Could minimize the residual in other norms
• For example we may have more confidence in some data, and want to be sure that their residual is lower
  – Attach a weight to each residual
    \[ \| r \|_w^2 = \sum_{i=1}^{m} w_i r_i^2 \]
• Or we may like the 1-norm or infinity norm better

\[
\| r \|_1 = \sum_{i=1}^{m} |r_i| \quad \| r \|_{\infty} = \max_{i} |r_i| 
\]
SVD and Pseudo-Inverse

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

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

$$x_r = \sum_{i=1}^{r} \frac{u_i^t b}{\sigma_i} v_i \quad \sigma_r > \varepsilon, \quad \sigma_{r+1} \leq \varepsilon$$

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