Computational Linear Algebra: Orthogonal Matrices, Gram-Schmidt Orthogonalization, and the QR Factorization

David Shuman

Orthogonality

Dot Products: As we know, the dot product between two vectors $\mathrm{v}, \mathrm{w} \in \mathbb{R}^n$ is given by

$$\mathrm{v} \cdot \mathrm{w} = \mathrm{v}^T \mathrm{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

The dot product satisfies

$$\mathrm{v} \cdot \mathrm{v} = ||\mathrm{v}||^2 \tag{1}$$

$$\mathrm{v} \cdot \mathrm{w} = ||\mathrm{v}||||\mathrm{w}|| \cos(\theta) \tag{2}$$

where θ is the angle between v and w (this is derived from the law of cosines)

Thus v and w are orthogonal if $\mathrm{v} \cdot \mathrm{w} = 0$

We say v and w are orthonormal if $\mathrm{v} \cdot \mathrm{w} = 0$ and $\mathrm{v} \cdot \mathrm{v} = \mathrm{w} \cdot \mathrm{w} = 1$

Orthogonal Matrices

- Orthogonal vectors are linearly independent
- If the columns of A are the orthonormal vectors $\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_n$, then

$$A^T A = I_n$$

- If A is square with *orthonormal* columns then

$$A^T A = I_n \qquad \Rightarrow \qquad \mathrm{A}^{-1} = \mathrm{A}^T$$

  - In this case A is called an orthogonal or orthonormal matrix.
    The inverse of an orthogonal matrix is its transpose!
- Solve: $\mathrm{A}x = \mathrm{b}$ with an orthogonal matrix A:

$$\mathrm{A}x = \mathrm{b} \quad \Rightarrow \quad \mathrm{A}^T \mathrm{A}x = \mathrm{A}^T \mathrm{b} \quad \Rightarrow \quad x = \mathrm{A}^T \mathrm{b}$$

Orthogonality is Nice: Length, Angles, and Conditioning

Let Q be an $n \times n$ orthogonal matrix and *behold*:

- For any $\mathrm{v} \in \mathbb{R}^n$

$$(Q\mathrm{v}) \cdot (Q\mathrm{v}) = (Q\mathrm{v})^T(Q\mathrm{v}) = (\mathrm{v}^T Q^T)(Q\mathrm{v}) = \mathrm{v}^T(Q^T Q)\mathrm{v} = \mathrm{v}^T\mathrm{v} = \mathrm{v} \cdot \mathrm{v}$$

$$\text{so } ||Q\mathrm{v}||^2 = ||\mathrm{v}||^2 \qquad \Rightarrow \qquad ||Q\mathrm{v}|| = ||\mathrm{v}||$$

  - Thus orthogonal matrices preserve length

- More generally, if $\mathrm{v}, \mathrm{w} \in \mathbb{R}^n$

$$(Q\mathrm{v}) \cdot (Q\mathrm{w}) = (Q\mathrm{v})^T(Q\mathrm{w}) = \mathrm{v}^T Q^T Q\mathrm{w} = \mathrm{v}^T\mathrm{w} = \mathrm{v} \cdot \mathrm{w}$$

  - Since $\mathrm{v} \cdot \mathrm{w} = ||\mathrm{v}||||\mathrm{w}|| \cos(\theta)$, we see that orthogonal matrices preserve angles between vectors

- $\mathrm{Cond}_2(Q) = ||Q||_2 \cdot ||Q^{-1}||_2 = \dfrac{\left(\max\limits_{||\mathrm{x}||_p=1} ||Q\mathrm{x}||_2\right)}{\left(\min\limits_{||\mathrm{x}||_p=1} ||Q\mathrm{x}||_2\right)} = 1$

  - Orthogonal matrices are as well-conditioned as possible

Orthogonality is Nice: Projections

- Let $\mathsf{q}_1, \mathsf{q}_2 \ldots, \mathsf{q}_n \in \mathbb{R}^m$ be a set of orthonormal vectors
- Let S be the subspace that they span; i.e., let $S = \text{col}(A)$ where

$$A = \begin{pmatrix} | & | & & | \\ \mathsf{q}_1 & \mathsf{q}_2 & \cdots & \mathsf{q}_n \\ | & | & & | \end{pmatrix}$$

- Suppose that $\mathsf{w} \in \mathbb{R}^m$ and we wish to find the orthogonal projection of w onto S
- We can use the orthogonal projection matrix

$$\begin{aligned} P\mathsf{w} &= A(A^T A)^{-1} A^T \mathsf{w} = AA^T \mathsf{w} \\ &= \mathsf{q}_1 \mathsf{q}_1^T \mathsf{w} + \mathsf{q}_2 \mathsf{q}_2^T \mathsf{w} + \cdots + \mathsf{q}_n \mathsf{q}_n^T \mathsf{w} \end{aligned}$$

  - Note that $\mathsf{q}_i \mathsf{q}_i^T$ is an outer product of vectors, *not* a dot product
  - Project independently onto each basis vector and add them up!

Example of a Projection with Orthogonal Vectors

$\mathbf{q}_1 = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})^T$ and $\mathbf{q}_2 = (\frac{2}{15}, \frac{2}{3}, -\frac{11}{15})^T$, and $A = \begin{pmatrix} \frac{1}{3} & \frac{2}{15} \\ \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{11}{15} \end{pmatrix}$

Then

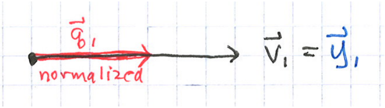
$$A^T A = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{15} & \frac{2}{3} & -\frac{11}{15} \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{15} \\ \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{11}{15} \end{pmatrix} = I_2$$

$$\begin{aligned} AA^T &= \begin{pmatrix} \frac{1}{3} & \frac{2}{15} \\ \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{11}{15} \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{15} & \frac{2}{3} & -\frac{11}{15} \end{pmatrix} = \frac{1}{225} \begin{pmatrix} 29 & 70 & 28 \\ 70 & 200 & -10 \\ 28 & -10 & 221 \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{pmatrix}}_{\mathbf{q}_1\mathbf{q}_1^T} + \underbrace{\begin{pmatrix} \frac{4}{225} & \frac{4}{45} & -\frac{22}{225} \\ \frac{4}{45} & \frac{4}{9} & -\frac{22}{45} \\ -\frac{22}{225} & -\frac{22}{45} & \frac{121}{225} \end{pmatrix}}_{\mathbf{q}_2\mathbf{q}_2^T} \end{aligned}$$

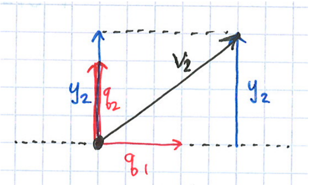
Gram-Schmidt: Orthogonalizes A Set of Vectors

Let $v_1, v_2 \ldots, v_n$ be a set of linearly independent vectors in $\mathbb{R}^m$. We will build up a set of vectors $q_1, q_2, \ldots, q_n$ which span the same space

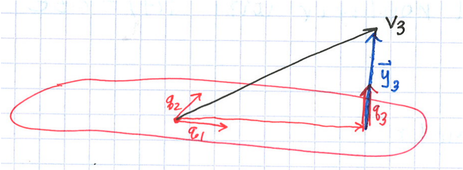
1. $y_1 = v_1$ and $q_1 = \frac{1}{||y_1||_2} y_1$



2. $$y_2 = v_2 - q_1 q_1^T v_2$$ $$q_2 = \frac{1}{||y_2||_2} y_2$$



3. $$y_3 = v_3 - (q_1 q_1^T v_3 + q_2 q_2^T v_3)$$ $$q_3 = \frac{1}{||y_3||_2} y_3$$ $$\vdots$$



Gram-Schmidt: Orthogonalizes A Set of Vectors

We can rewrite the equations on the previous page to represent the v_i's in terms of the q_j's as follows:

$$\begin{aligned}
\mathrm{v}_1 &= ||\mathrm{y}_1||\mathrm{q}_1 \\
\mathrm{v}_2 &= ||\mathrm{y}_2||\mathrm{q}_2 + (\mathrm{q}_1 \cdot \mathrm{v}_2)\mathrm{q}_1 \\
\mathrm{v}_3 &= ||\mathrm{y}_3||\mathrm{q}_3 + (\mathrm{q}_2 \cdot \mathrm{v}_3)\mathrm{q}_2 + (\mathrm{q}_1 \cdot \mathrm{v}_3)\mathrm{q}_1 \\
&\ \vdots
\end{aligned}$$

or, as matrices,

$$\begin{pmatrix} | & | & | \\ \mathrm{v}_1 & \mathrm{v}_2 & \mathrm{v}_3 \\ | & | & | \end{pmatrix} = \underbrace{\begin{pmatrix} | & | & | \\ \mathrm{q}_1 & \mathrm{q}_2 & \mathrm{q}_3 \\ | & | & | \end{pmatrix}}_{Q} \underbrace{\begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ 0 & r_{2,2} & r_{2,3} \\ 0 & 0 & r_{3,3} \end{pmatrix}}_{R}$$

with

$$r_{i,i} = ||\mathrm{y}_i|| \qquad \text{and} \qquad r_{i,j} = \mathrm{q}_i \cdot \mathrm{v}_j$$

QR Decomposition (Reduced Form)

The QR Decomposition is a matrix decomposition that writes any $m \times n$ matrix A with $m \geq n$ as a product

$$A = QR$$

where

- Q is an $m \times r$ matrix with orthonormal columns, where r is the number of linearly independent columns of A (i.e., $r = n$ if A is full rank)
- R is an $r \times n$ matrix which is upper triangular if $r = n$, or the top portion of an upper triangular matrix if $r < n$
- The columns of Q span the same space as the columns of A
- The matrix R gives the change of basis between the vectors in Q and the vectors in A
- It is unique up to some sign changes, so if we require $R_{ii} \geq 0$, it is unique
- If the columns of A are independent, then $R_{ii} \neq 0$
- On the other hand, if column a_j can be written as a linear combination of columns $a_1, \ldots, a_{j-1}$, then $R_{jj} = 0$

QR Decomposition: Examples

- Example 1 (square)

$$\begin{pmatrix} 2 & 10 & 19 \\ 3 & 8 & -3 \\ 6 & 9 & -13 \end{pmatrix} = \begin{pmatrix} \frac{2}{7} & \frac{6}{7} & \frac{3}{7} \\ \frac{3}{7} & \frac{2}{7} & -\frac{6}{7} \\ \frac{6}{7} & -\frac{3}{7} & \frac{2}{7} \end{pmatrix} \begin{pmatrix} 7 & 14 & -7 \\ 0 & 7 & 21 \\ 0 & 0 & 7 \end{pmatrix}$$

- Example 2 (linearly dependent columns)

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 2 & -2 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{2}{3} \end{pmatrix} \begin{pmatrix} 3 & 0 & 1 \\ 0 & 3 & 1 \end{pmatrix}$$

- Example 3 (tall)

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{1}{2} & \frac{1}{2\sqrt{3}} & -\sqrt{\frac{2}{3}} \\ \frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{\sqrt{3}} \\ 0 & 0 & \sqrt{\frac{2}{3}} \end{pmatrix}$$

QR Decomposition (Full Decomposition)

$$m\left\{\overbrace{\begin{bmatrix} \quad A \quad \end{bmatrix}}^{n}\right. = \underbrace{\begin{bmatrix} \overbrace{\quad Q \quad}^{n} & \overbrace{\quad \hat{Q} \quad}^{m-n} \end{bmatrix}}_{\bar{Q}} \underbrace{\overbrace{\begin{bmatrix} R \\ \hline 0 \end{bmatrix}}^{n}}_{\bar{R}} \begin{array}{l} \}\, n \\ \}\, m-n \end{array}$$

The additional $m - n$ columns of $\bar{Q}$ give a basis for the orthogonal complement of the column space of A, which is equal to the null space of $A^\top$ (see also the four fundamental subspaces handout)

Computing the QR Decomposition

- Classical Gram-Schmidt
  - Computational cost: $O(2mn^2)$ multiplications and additions
  - Can also be viewed as *triangular orthogonalization*: $AR_1R_2\dots R_n = Q$
- Modified Gram-Schmidt
  - Instead of projecting the next column onto each previously computed basis vector and subtracting those away to find the new vector, subtract away the first projection before computing the second projection, etc.
  - Same computational cost, but more stable to rounding errors
- Householder reflections
  - Method of choice in most implementations
  - Householder reflectors are super cool! May revisit after the midterm
  - Can be viewed as *orthogonal triangularization* $Q_n \dots Q_2Q_1A = R$
  - Computational cost: $O(2mn^2 - \frac{2}{3}n^3)$ multiplications and additions
  - If you are only interested in computing matrix-vector products like $Q^\top b$ (as in least squares), you can do even better in terms of computation and memory because you don't actually have to form and store Q
  - Also more stable to rounding errors

Three Uses of the QR Decomposition

1. Solving $\mathrm{A}x = \mathrm{b}$ (if A is square)

$$\begin{aligned} \mathrm{A}x &= \mathrm{b} \\ \mathrm{QR}x &= \mathrm{b} \\ \mathrm{Q}^T\mathrm{QR}x &= Q^T\mathrm{b} \\ \mathrm{R}x &= \mathrm{Q}^T\mathrm{b} \end{aligned}$$

- This is upper triangular, so back substitution works
- The hard work is in computing the QR decomposition
- Three times more computations than the LU decomposition, so you wouldn't usually solve $Ax = b$ this way

Three Uses of the QR Decomposition

2. Least squares for $\mathrm{Ax} = \mathrm{b}$: If $\bar{x}$ minimizes

$$||A\mathrm{x} - \mathrm{b}|| = ||QR\mathrm{x} - \mathrm{b}|| = ||\bar{Q}\bar{R}\mathrm{x} - \mathrm{b}||$$

   Then since multiplication by an orthogonal matrix Q^T does not change the length, we can minimize

$$||A\mathrm{x} - \mathrm{b}||^2 = ||\bar{Q}\bar{R}\mathrm{x} - \mathrm{b}||^2 = ||\bar{Q}^T(\bar{Q}\bar{R}\mathrm{x} - \mathrm{b})||^2 = ||\bar{R}\mathrm{x} - \bar{Q}^T\mathrm{b}||^2$$
$$= ||R\mathrm{x} - Q^T\mathrm{b}||^2 + ||\hat{Q}^T\mathrm{b}||^2$$

   - To minimize this, we can solve $\mathrm{Rx} = \mathrm{Q}^T\mathrm{b}$ to make the first term above equal to 0
   - The second part of the error term does not depend at all on our choice of x
   - Complexity dominated by QR factorization: $O(2mn^2)$
   - Solving same problem with Cholesky: $O(mn^2 + \frac{1}{3}n^3)$ (roughly half for $m >> n$)
   - However, QR is more accurate (less affected by rounding errors) and is therefore often used in small to medium size problems
   - For large, sparse matrices, Cholesky does a better job at preserving and exploiting the sparsity

3. Computing eigenvalues and eigenvectors — later