Mostafa Sabri

Lecture 21

6.4 The Gram-Schmidt Process

6.5 Least Square Problems

جامعة نيويورك أبوظبي
NYU ABU DHABI

The Gram-Schmidt Process

Suppose $W = \mathrm{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent in $\mathbb{R}^n$, and we want to construct an orthogonal basis for W. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ be this basis. We can take the first vector in $\mathcal{B}$ to be $\mathbf{x}_1$ itself, i.e. $\mathbf{v}_1 = \mathbf{x}_1$. For the second vector, we need it

(i) To be orthogonal to $\mathbf{v}_1 = \mathbf{x}_1$,

(ii) To be a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$.

Recalling Lec20, we take $\mathbf{v}_2 = \mathbf{x}_2 - \widehat{\mathbf{x}}_2$, where $\widehat{\mathbf{x}}_2$ is the orthogonal projection of $\mathbf{x}_2$ onto $\mathbf{x}_1$. Then $\mathbf{v}_2 \perp \mathbf{x}_1$ and $\mathbf{v}_2 = \mathbf{x}_2 - \alpha\mathbf{x}_1$ satisfies (ii) as well.

The Gram-Schmidt Process is simply an iteration of this.

The Gram-Schmidt Process

Thm (Gram-Schmidt): Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ be a basis for a nonzero subspace W of $\mathbb{R}^n$. Define

$$\mathbf{v}_1 = \mathbf{x}_1\,,$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1\,,$$

$$\vdots$$

$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$

(i) Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W.

(ii) Moreover, for any $1 \leq k \leq p$, we have

$$\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}\,. \tag{1}$$

The Gram-Schmidt Process

Proof. The proof is by induction. If $p = 1$ the statement is trivial. Now suppose (i)-(ii) are true for $p = k$. Let $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ and define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1} \tag{2}$$

From Lec20, we get that $\mathbf{v}_{k+1} \in W_k^{\perp}$. Also, $\text{proj}_{W_k} \mathbf{x}_{k+1} \in W_k \subset W_{k+1}$, so $\mathbf{v}_{k+1} \in W_{k+1}$ as W_{k+1} is a subspace. Finally $\mathbf{v}_{k+1} \neq \mathbf{0}$ since $\mathbf{x}_{k+1} \notin W_k$. Using our hypothesis that the theorem holds for $p = k$, and in view of (1), we deduce that $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal set in W_{k+1}, which has dimension $k + 1$. Since it is in particular linearly independent, it must be a basis for W_{k+1} by the basis theorem. This proves (ii) for $p = k + 1$, and also (i) since $W = W_{k+1}$ if $p = k + 1$. $\square$

Orthonormal Bases

It follows that any nonzero subspace W of $\mathbb{R}^n$ has an orthogonal basis.

This implies that any such W also has an orthonormal basis, simply by normalizing the vectors $\mathbf{v}_k$ of the orthogonal basis.

Find an orthonormal basis for $W = \text{Span}\{(1, 2, -1, 0), (2, 2, 0, 1), (1, 1, -1, 0)\}$.

Computer algorithms can apply the Gram-Schmidt process to the columns of a matrix A. The result is a *QR factorization for A.* This can be useful for solving equations.

QR Factorization

Thm (QR Factorization): Let A be an $m \times n$ matrix with linearly independent columns. Then A can be factored as $A = QR$, where

- Q is an $m \times n$ matrix whose columns form an orthonormal basis of $\mathrm{Col}(A)$,
- R is an $n \times n$ upper triangular invertible matrix with positive diagonal entries.

Proof. The columns $\mathcal{B} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ form a basis of $\mathrm{Col}(A)$. Apply the Gram-Schmidt process to $\mathcal{B}$ and normalize the resulting vectors to obtain an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of $\mathrm{Col}(A)$. Let $Q = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Since $\mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} = \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, there are

QR Factorization

some $r_{ik} \in \mathbb{R}$ such that

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_n\,. \qquad (3)$$

We may assume that $r_{kk} \geq 0$ by multiplying both r_{kk} and $\mathbf{u}_k$ by -1 if necessary.

Equation (3) says that $\mathbf{x}_k = Q\mathbf{r}_k$, where $\mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. If $R = \begin{bmatrix} \mathbf{r}_1 & \cdots & \mathbf{r}_n \end{bmatrix}$, then R is clearly upper triangular and $A = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} Q\mathbf{r}_1 & \cdots & Q\mathbf{r}_n \end{bmatrix} = QR$.

QR Factorization

To see that R is invertible, suppose $R\mathbf{x} = \mathbf{0}$. Then $A\mathbf{x} = QR\mathbf{x} = Q\mathbf{0} = \mathbf{0}$. As A has linearly independent columns, the only solution to $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$ (Lec3). It follows that R is invertible (Lec7).

Finally, $\det(R) = r_{11}r_{22}\cdots r_{nn}$ and $r_{kk} \geq 0$. Since R is invertible, we must have $r_{kk} > 0$ for all k. $\square$

Find a QR factorization for $A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$.

Least-Squares Problems

Inconsistent systems often arise in applications. If $A\mathbf{x} = \mathbf{b}$ has no solution, we can try to search for the best approximation $\widehat{\mathbf{x}}$ of a solution, i.e. try to find $\widehat{\mathbf{x}}$ such that $\|A\widehat{\mathbf{x}} - \mathbf{b}\|$ is as small as possible.

If A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$, a *least-squares solution* of $A\mathbf{x} = \mathbf{b}$ is an $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|\mathbf{b} - A\widehat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \qquad \forall \mathbf{x} \in \mathbb{R}^n.$$

Since $A\mathbf{x} = \sum_{i=1}^{n} x_i \boldsymbol{\alpha}_i \in \mathrm{Col}(A)$, we can simply take $\widehat{\mathbf{x}} = \mathbf{x}_0$, where $A\mathbf{x}_0 = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$, by Lec20. If $\mathbf{b} \in \mathrm{Col}(A)$, then $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}_0$ (Lec2), and we can take $\widehat{\mathbf{x}} = \mathbf{x}_0$.

Least-Squares Problems

Thm: The set of least-square solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the *normal equations* $A^T A\mathbf{x} = A^T\mathbf{b}$.

Proof. Let $\widehat{\mathbf{b}} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$. The set of least-squares solutions is the set of solutions $\mathbf{y}$ of $A\mathbf{y} = \widehat{\mathbf{b}}$.

Since $\mathbf{b} - \widehat{\mathbf{b}} \in \mathrm{Col}(A)^\perp = \mathrm{Nul}(A^T)$, if $A\mathbf{y} = \widehat{\mathbf{b}}$, then $A^T A\mathbf{y} = A^T\widehat{\mathbf{b}} = A^T(\mathbf{b} - (\mathbf{b} - \widehat{\mathbf{b}})) = A^T\mathbf{b}$.

Conversely, if $A^T A\mathbf{y} = A^T\mathbf{b}$, then $A^T(\mathbf{b} - A\mathbf{y}) = \mathbf{0}$. So $\mathbf{b} = A\mathbf{y} + (\mathbf{b} - A\mathbf{y})$, and $\mathbf{b} - A\mathbf{y} \in \mathrm{Nul}(A^T) = \mathrm{Col}(A)^\perp$, while $A\mathbf{y} \in \mathrm{Col}(A)$. By uniqueness of orthogonal decomposition, we must have $A\mathbf{y} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) = \widehat{\mathbf{b}}$. $\square$

Least-Squares Problems

Find a least-squares solution to the problem $A\mathbf{x} = \mathbf{b}$ if

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}.$$

The terminology "least-squares" is due to the fact that we are trying to find an $\mathbf{x}$ minimizing $\sum_{i=1}^{n} |b_i - (A\mathbf{x})_i|^2$, instead of one minimizing $\sum_{i=1}^{n} |b_i - (A\mathbf{x})_i|$ for example.

Find the orthogonal projection of $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ onto the column space of $A = \begin{bmatrix} 0 & 2 \\ 3 & 0 \\ 1 & 0 \end{bmatrix}$.

Unique least-squares solution

Thm: Let A be an $m \times n$ matrix. The following statements are equivalent.

(1) The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b} \in \mathbb{R}^m$.

(2) The columns of A are linearly independent.

(3) The matrix $A^T A$ is invertible.

If any of these statements holds, the unique solution is $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.

Proof. (1) $\Longrightarrow$ (3). If (1) holds then $A^T A\mathbf{x} = A^T\mathbf{b}$ has a unique solution for any $\mathbf{b}$. Taking $\mathbf{b} = \mathbf{0}$, we get that $A^T A\mathbf{x} = \mathbf{0}$ has a unique solution, which must be the trivial solution. So (3) holds (Lec7).

Unique least-squares solution

(3) $\Longrightarrow$ (1) Conversely, if (3) holds, then $A^TA\mathbf{x} = A^T\mathbf{b}$ has the unique solution $\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}$ so (1) holds.

(2) $\Longleftrightarrow$ (3) We first prove that $\mathrm{Nul}(A) = \mathrm{Nul}(A^TA)$. In fact, if $A\mathbf{x} = \mathbf{0}$ then $A^TA\mathbf{x} = \mathbf{0}$. This shows $\mathrm{Nul}(A) \subseteq \mathrm{Nul}(A^TA)$. Conversely, if $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x}^T(A^TA\mathbf{x}) = 0$, so $(\mathbf{x}^TA^T)(A\mathbf{x}) = 0$, so $(A\mathbf{x})^T(A\mathbf{x}) = 0$ and thus $A\mathbf{x} = 0$. Thus, $\mathrm{Nul}(A) = \mathrm{Nul}(A^TA)$.

It follows (Lec3) that the columns of A are linearly independent iff the columns of A^TA are linearly independent, which occurs iff A^TA is invertible (Lec7). $\square$

If $\widehat{\mathbf{x}}$ is a least-squares solution, we call $\|\mathbf{b} - A\widehat{\mathbf{x}}\|$ the *least square error.*

QR Factorizations and Least-Square Solutions

Find the least-squares errors in the previous examples.

The following theorem gives a method which is often more reliable for computer calculations.

Thm: Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization for A. Then for each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

Proof. We already know the solution is unique. Let $\widehat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$. Then $A\widehat{\mathbf{x}} = QQ^T\mathbf{b}$. As Q has orthonormal columns, then $QQ^T\mathbf{b} = \text{proj}_{\text{Col}(A)}(\mathbf{b})$ by Lec20. So $\hat{\mathbf{x}}$ is a least-squares solution. $\square$

QR Factorizations and Least-Square Solutions

Find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ if

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 4 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$