Orthogonalization (QR) & LS approximation

Learning objectives: get familiar with an orthogonalization process, the Gram-Schmid process in particular, and the key components

▷ Each step, after the initial step, involves the orthogonal decomposition of the "next" vector via orthogonal projections
▷ The invariance in preserving a subspace sequence across successive steps
▷ The upper triangular structure of the projection coefficient matrix (QR factorization)
▷ The algorithmic complexity in terms of the number of inner products

and the main application to LS approximations, analytically as well as numerically.

Preliminary: orthogonal projections & decomposition

Assume an inner product space. Denote by $\| \cdot \|$ the 2-norm induced by the inner product.

The base case. Let $q$ be a unit-length vector, $\|q\| = 1$. we have the orthogonal decomposition of $v$

$$v = P(v) + P^\perp(v), \quad \|v\|^2 = \|P(v)\|^2 + \|P^\perp(v)\|^2, \quad v \in V. \quad (1)$$

where the two (complementary) orthogonal projections are

$$P(v) = q < q, v >, \quad P^\perp = v - P(v), \quad v \in V. \quad (2)$$

The projectors $P$ and $P^\perp$ are linear mappings.

An extension to multiple-dimensions. Let $Q = \{q_1, \cdots, q_j\}$ be a set of orthonormal vectors, $< q_i, q_j > = \delta_{ij}$. We have the orthogonal decomposition

$$v = P_i(v) + P^\perp_j(v), \quad \|v\|^2 = \|P_i(v)\|^2 + \|P^\perp_j(v)\|^2 \quad v \in V. \quad (3)$$

where the orthogonal projectors and projections are

$$P_i(v) = \sum_{i=1}^j q_j < q_j, v >, \quad P^\perp_j(v) = v - P_i(v). \quad v \in V. \quad (4)$$

The projectors are linear mappings.
The Gram-Schmidt process

Input: a set of linearly independent vectors \( \{v_1, v_2, \cdots, v_k\} \) in an inner product space \((V, F, \langle \cdot, \cdot \rangle)\).

Output: a set of orthonormal vectors \( \{q_1, q_2, \cdots, q_k\} \), with the property of preserving the subspace sequence

\[
\text{span}(q_1, \cdots, q_j) = \text{span}(v_1, \cdots, v_j), \quad j \leq k. \tag{5}
\]

1. normalize the input vectors: \( v_j := v_j / \|v_j\|, 1 \leq j \leq k \)
2. initial setting: \( q_1 := v_1 \)
3. For \( j = 1 : k - 1 \) % successive steps
   let \( P_j \) be the orthogonal projector to \( Q_j = \text{span}(q_1, \cdots, q_j) \)
   get the orthogonal residual of the next input vector \( v_{j+1} \) to \( Q_j \),
   \[
   r_{j+1} := v_{j+1} - P_j(v_{j+1}), \quad \text{by the decomposition:} v_{j+1} = P_j(v_{j+1}) + r_{j+1} \tag{6}
   \]
   \( q_{j+1} := r_{j+1} / \|r_{j+1}\| \)

The next output vector is the orthogonal complement normalized.

Preserving the subspace sequence

- The property (5) is the invariance across the successive G-S steps.
- The property is well captured in the QR factorization of the input vectors expressed as follows,

\[
[v_1, v_2, \cdots, v_k] D = [q_1, q_2, \cdots, q_k] R, \tag{7}
\]

where \( D = \text{diag}(\|v_1\|^{-1}, \cdots, \|v_k\|^{-2}) \) is the scaling matrix, and \( R \) is the matrix of projection coefficients,

\[
R(i, j) = \begin{cases} 
\langle q_i, v_j \rangle, & i < j \\
\left(1 - \sum_{i=1}^{j-1} \langle q_i, v_j \rangle\right)^{1/2}, & i = j \\
0, & i > j 
\end{cases} \tag{8}
\]

Matrix \( R \) is upper triangular and non-singular. The upper triangular structure of \( R \) indicates clearly that \( v_j \in \text{span}(q_1, \cdots, q_j) \).

- For each column \( j \) of \( R \), the squared elements sum to 1. The orthogonal decomposion of \( v_j \) with respect to \( Q_i = \text{span}(q_1, \cdots, q_j) \), for any \( i > j \), is determined by \( R(1 : i, j) \) and \( R(i + i : j, j) \).
Rank-revealing orthogonalization

Input: a set of \( p \) non-zero vectors \( \{v_1, v_2, \cdots, v_p\} \), not necessarily linearly independent

Output: a set of \( k \) orthonormal vectors \( \{q_1, q_2, \cdots, q_k\} \), \( k \leq p \), such that

\[
\langle q_i, q_j \rangle = \delta_{ij}, \quad \text{span}(q_1, \cdots, q_k) = \text{span}(v_1, \cdots, v_p)
\]

The following is a conceptual description, which can be used analytically not numerically.

1. normalize the input vectors: \( v_j := v_j / \|v_j\|, 1 \leq j \leq p \)
2. initial setting: \( q_1 := v_1, k := 1 \)
3. For \( j = 1 : p - 1 \) % successive steps
    - let \( P_k \) be the projector to \( \text{span}(q_1, \cdots, q_k) \)
    - set \( u_{k+1} := v_{j+1} \) and decompose \( u_{k+1} \),
    - \[ r_{k+1} := u_{k+1} - P_k(u_{k+1}), \quad (9) \]
    - If \( r_{k+1} \neq 0 \) (*) \(^1\), set
      \[
      q_{k+1} := r_{k+1} / \|r_{k+1}\|, \quad k := k + 1
      \]
    - Else move \( v_{j+1} \) to the end of the input vector sequence.

Rank-revealing QR factorization

\[
[v_1, v_2, \cdots, v_p] D \Pi = [q_1, q_2, \cdots, q_k] R_{k \times p} \quad (10)
\]

where the \( k \times k \) leading submatrix of \( R \) is nonsingular and \( \Pi \) is a permutation matrix. The first \( k \) vectors by the recording are linearly independent.

\(^1\)This mathematical condition must be converted to a numerical one in computation. This conversion will trigger the inspection and analysis of more fundamental numerical issues with the GS-QR process.
Relation to the LS approximation theory & methods

Case 1: orthogonal vectors in $\mathbb{R}^n$

Consider the inner product space with $V = \mathbb{R}^n$, $F = \mathbb{R}$ and

$$< u, v > = \sum_{i,j=1}^{n} u(i)v(j)M(i,j)$$

(11)

where $M$ a symmetric, positive definite matrix.

1. $M = I$
2. $M$ is diagonal, $M \neq I$
3. $M$ is not diagonal

Case 2: orthogonal polynomial families

Consider the inner product space with $V = P_{n-1}[-1,1]$, $F = \mathbb{R}$ and

$$< p, q > = \int_{-1}^{1} p(x)q(x)w(x)dt. $$

(12)

where $w(x) > 0$ on $(-1,1)$ is a weight function, or $w(x)dx$ is the infinitesimal measure.

1. $w(x) = 1$ Legendre polynomials $P_j(x)$ (The notion is by convention). They were introduced initially for the expansion of the Newtonian potential, a.k.a., the multipole expansion for the solution of time-independent 3D Laplace equation. These polynomials are used recently in recurrent networks as architectural units.

2. $w(x) = 1/\sqrt{(1-x)(1+x)}$ Chebyshev polynomials $T_j(x)$ of the first kind (a connection between polynomials and trigonometric polynomials)

They have many roles in analysis as well as in approximation. For example, they are used in the multipole expansions for the solution of 2D Laplace equations. They have a unique role in connecting polynomials and trigonometric polynomials.

3. $w(x) = (1-x)^\alpha(1+x)^\beta$ for Jacobi polynomials with parameters $(\alpha, \beta)$, a.k.a., hypergeometric polynomials. It is the Legendre family when $\alpha = \beta = 0$, and it is the Chebyshev family when $\alpha = \beta = -1/2$.

Exercise

Get familiar with the utilities of the QR factorization

1. Let $\{v_1, v_2, \ldots, v_n\}$ be a set of vectors in the space $\mathbb{R}^n$, $n > 1$. Then, the dimension of span($v_1, \ldots, v_n$) detected/determined by a rank-revealing orthogonalization process in exact arithmetic operations with respect to an inner product of (??) is mathematically the same as that with respect to another inner product of (??). Numerically, with arithmetic operations in finite precision, the numbers may differ.

2. Let $\{v_1, v_2, \ldots, v_n\}$ be a set of linearly independent vectors in $\mathbb{R}^n$. Let $\{q_1, \ldots, q_n\}$ \{\tilde{q}_1, \ldots, \tilde{q}_n\} be two orthogonal basis sets for the space span($v_1, \ldots, v_n$) with respect to two different inner products of (??), and span($q_1, \ldots, q_j$) = span($\tilde{q}_1, \ldots, \tilde{q}_j$) = span($v_1, \ldots, v_j$), $j \leq n$. Then,

$$[\tilde{q}_1, \ldots, \tilde{q}_j] = [q_1, \ldots, q_n] R_{n \times n}$$

where matrix $R_{n \times n}$ is upper triangular, not diagonal.
3. Let \( \{q_0(t), q_1(t), \ldots, q_n(t)\} \) and \( \{\tilde{q}_0(t), \tilde{q}_1(t), \ldots, \tilde{q}_n(t)\} \) be two orthogonal polynomial sets, with respect to two different inner products of (12), and

\[
\text{span}(q_0(t), \ldots, q_j(t)) = \text{span}(\tilde{q}_0(t), \ldots, \tilde{q}_j(t)) = \text{span}(1, t, \ldots, t^j), \quad j \leq n
\]

Then, the two families are related as follows,

\[
[\tilde{q}_0(t), \ldots, \tilde{q}_j(t)] = [q_0(t), \ldots, q_j(t)] R
\]

where matrix \( R \) is non-singular, upper triangular, not diagonal.

4. Let \( \{x_i\}_{i=1}^m \) be \( m \) knot nodes in \([-1, 1]\). Let \( V_{m \times n} = V(\{x_i\}; \{1, x, \ldots, x^{n-1}\}) \) be the Vandermonde matrix. Let \( Q_{m \times n} = Q(\{x_i\}; q_0(x), q_1(x), \ldots, q_{n-1}(x)) \) be the matrix by an orthogonal polynomial set on the same nodes. Then,

(a) there is a non-singular upper triangular matrix \( R_{n \times n} \) such that

\[
V_{m \times n} = Q_{m \times n} R_{n \times n}.
\]

Consequently, \( V_{m \times n} \) and \( Q_{m \times n} \) have the same column rank.

(b) it can be proven that if the knots are distinctive and \( m \geq n \), then \( V_{n \times n} \) is of full column rank. However, the columns of \( Q_{n \times n} \) are not necessarily mutually orthogonal except at certain set of knots.

(c) For exact polynomial fitting (interpolation), we get the same polynomial interpolant with \( V_{n \times n} \) or \( Q_{m \times n} \) if the knots are distinctive and \( m \geq n \).

(d) For polynomial fitting with the least squared residual, up to a fixed degree, we get the same polynomial interpolant with \( V_{n \times n} \) or \( Q_{m \times n} \) if the knots are distinctive and \( m \geq n \).
Advanced exercise

Recurrence among a walk sequence

Consider the inner product space $\mathbb{R}^n$ over the field $\mathbb{F} = \mathbb{R}$ and equipped with an inner product of (11). Let $A$ be a symmetric matrix, $A \in \mathbb{R}^n$. It may represent an undirected network. Given a nonzero vector $v, v \in \mathbb{R}^n$. Consider the walk sequence

$$v, Av, A^2v, \cdots, A^k v, \cdots$$

and the sequence of the successively spanned subspaces $\{V_j \mid j \in \mathbb{N}\}$ with

$$V_j = \text{span}(v, Av, A^2v, \cdots, A^j v),$$

We may use an orthogonalization process to determine the dimension as $j$ increases.

- **Initial:** let $q_0 = v/\|v\|$ and $q_1$ be the first two orthogonal vectors $\text{span}(q_0, q_1) = \text{span}(v, Av)$.
- **For** $j = 1, 2, \cdots$, until $\|r_{j+1}\| < \tau$, $\tau$ is a numerical threshold
  let $P_j$ be the projector to $\text{span}(q_1, \cdots, q_j)$
  decompose the “next” vector $A q_j$ by $P_j$,

$$r_{j+1} := A q_j - P_j(A q_j)$$

$$q_{j+1} := r_{j+1}/\|r_{j+1}\|$$

(13)

where it is straightforward to verify that

$$P_j(A, q_j) = q_{j-1} < q_{j-1}, A q_j > + q_j < q_j, A q_j >.$$  
(14)

By (14), the orthogonal vectors satisfy a three-term recurrence relation. Any new vector depends only the most recent two, instead of all the previous ones. In other words, the total number of inner products up to step $j$ is $O(j)$ instead of $O(j^2)$.

**Exercise**

1. Let $\ell$ be large enough, say, $\ell > n$. Then, $V_k = V_\ell$, $k > \ell$.
2. Let $r$ be the smallest integer so that $V_{k'} = V_r$, $k' > r$. Then, $r \leq \text{rank}(A) - 1$.
3. Give a condition under which $r < \text{rank}(A) - 1$.
4. The invariance across the successive steps in preserving the subspace sequence

$$\text{span}(q_0, q_1, \cdots, q_j) = \text{span}(v, Av, \cdots, A^j v), \quad j < r.$$  
(15)

5. Suppose $r > 3$. Let $T$ be representation of matrix $A$ restricted to the subspace $V_r$ in the orthogonal basis,

$$A[q_0, q_1, \cdots, q_r] = [q_0, q_1, \cdots, q_r] T.$$  
(16)

Then $T$ is symmetric and tridiagonal. The above process is known as the Lanczos process for transforming a symmetric matrix to a tridiagonal form. Among other purposes, it is used for model reduction for LS approximation, or regulated iteration for obtaining LS solutions.

Without the symmetry, the above procedure works with all projections to all previous orthogonal vectors, and results in the reduction to the Hessenberg form. The process is known as the Arnoldi method, used for model reduction or regulated iteration in LS approximation.
Recurrence among orthogonal polynomials

The polynomial subspaces have the following recurrence relationship,

\[
\text{span}(1, t, \ldots, t^j, t^{j+1}) = c + t \cdot \text{span}(1, t, \ldots, t^j)
\]  

We can apply the Gram-Schmidt procedure in the following recurrence manner to obtain a family of orthogonal polynomials with respect to an inner product of (12).

- **Initial**: let \( q_0 \) and \( q_1 \) be the first two orthogonal polynomials so that \( \text{span}(q_0, q_2) = \text{span}(1, t) \)

- **For \( j = 1, 1, \cdots \)** % successive steps
  - let \( P_j \) be the projector to \( \text{span}(q_1, \cdots, q_j) \)
  - decompose the “next” vector \( t \cdot q_j \) by \( P_j \),
    \[
    r_{j+1} := t \cdot q_j - P_j(t \cdot q_j) \\
    q_{j+1} := r_{j+1}/\|r_{j+1}\|
    \]

where it is straightforward to verify that

\[
P_j(t \cdot q_j) = q_{j-1} < q_{j-1}, t \cdot q_j > + q_j < q_j, t \cdot q_j >. 
\]

One can verify that the invariance

\[
\text{span}(q_0, q_1, \cdots, q_j) = \text{span}(1, t, \cdots, t^j). 
\]

In general, we have the three-term recurrence relation among the orthogonal polynomials

\[
\gamma_{j+1} q_{j+1}(x) = \alpha_{j+1} x q_j(x) + \beta_{j+1} q_{j-1}(x) 
\]

where the recurrence coefficients \( \{\alpha_j, \beta_j, \gamma_j\} \) are specific to the inner product. This is very useful for deriving the differential and integral equations among the polynomials. For convenience in expression, the orthogonal polynomials are not necessarily normalized.

In particular,

- **Legendre polynomials:**
  \[
  (j + 1)P_{j+1}(x) = (2j + 1)x P_j(x) - j P_{j-1}(x) 
  \]
  with \( P_0(x) = 1 \) and \( P_1(x) = x \).

- **Chebyshev polynomials:**
  \[
  T_{j+1}(x) = 2x T_j(x) - T_{j-1}(x) 
  \]
  with \( T_0(x) = 1 \) and \( T_1(x) = x \).